Vol. 35, No. 4 July, 1928

PSYCHOLOGICAL REVIEW PUBLICATIONS

Psychological Review

EDITED BY

HOWARD C. WARREN, Princeton University
MADISON BENTLEY, University of Illinois (*J. of Exper. Psychol.*)
S. W. FERNBERGER, Univ. of Pennsylvania (*Bulletin*)
W. S. HUNTER, Clark University (*Index*), and
RAYMOND DODGE, Yale University (*Monographs*)
HERBERT S. LANGFELD, Princeton University, *Business Editor*

ADVISORY EDITORS

R. P. ANGIER, Yale University; MARY W. CALKINS, Wellesley College; JOSEPH JASTROW, University of Wisconsin; C. H. JUDD, University of Chicago; ADOLF MEYER, Johns Hopkins University; W. B. PILLSBURY, Univ. of Michigan; C. E. SEASHORE, University of Iowa; G. M. STRATTON, Univ. of California; MARGARET F. WASHBURN, Vassar College; JOHN B. WATSON, New York; R. S. WOODWORTH, Columbia University.

CONTENTS

The Theory of Factors, II: Stuart C. Dodd, 261.

A Suggestive Review of Gestalt Psychology: H. H. Hsiao, 280.

The Gestalt Enigma: Horace G. Wyatt, 298.

An Eclectic Theory of Vision: Gilbert J. Rich, 311.

Are There Any Native Emotions?: J. F. Dashiell, 319.

Some Fallacies Underlying the Use of Psychological 'Tests': H. M. Johnson, 328.

Discussion:

On the Presuppositions of Character Testing: Herbert Woodrow, 338.

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Psychological Review Publications

OF THE

American Psychological Association

EDITED BY

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
RAYMOND DODGE, YALE UNIVERSITY (*Monographs*)
SAMUEL W. FERNBERGER, UNIV. OF PENNSYLVANIA (*Bulletin*)
W. S. HUNTER, CLARK UNIVERSITY (*Index*)
MADISON BENTLEY, UNIVERSITY OF ILLINOIS (*J. of Exp. Psych.*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY, *Business Editor*

WITH THE COOPERATION OF
MANY DISTINGUISHED PSYCHOLOGISTS

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bimonthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 480 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly, the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bimonthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 500 pages.

PSYCHOLOGICAL INDEX

Is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in May, and may be subscribed for in connection with the periodicals above, or purchased Maarately.

PSYCHOLOGICAL MONOGRAPHS

consists of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

ANNUAL SUBSCRIPTION RATES

Review: $5.50 (Foreign, $5.75)
Journal: $6.00 (Foreign, $6.25)
Bulletin: $6.00 (Foreign, $6.25)
Monographs: $6.00 per volume (Foreign, $6.30)
Psychological Index: $4.00

Current Numbers: Review or Journal, $1.00; Bulletin, $.60

Review and Bulletin, $10.00 (Foreign, $10.50)
Journal and Bulletin, $11.00 (Foreign, $11.50)
Review and Journal, $10.00 (Foreign, $10.50)
Review, Bulletin, and Journal, $15.00 (Foreign, $15.75)
Review, Bulletin, Journal, and Index, $18.00 (Foreign, $18.75)

Subscriptions, orders, and business communications may be sent direct to

PSYCHOLOGICAL REVIEW COMPANY,
PRINCETON, N. J.

VOL. 35, No. 4 July, 1928

THE PSYCHOLOGICAL REVIEW

THE THEORY OF FACTORS, II

BY STUART C. DODD

National Research Council Fellow 1926–27

In the Psychological Review for 1920, Spearman and Thomson restated their positions in a terminology less technical than that used in their previous articles. In his reviewing statement (**18**), Spearman commented with surprise on the little use that psychologists had made of the theory of '*g*,' and thought that this lack of use was due to ignorance and biased methods of thinking. (The present reviewer feels that psychologists often object to so simple an analysis as a dichotomus '*g*' and '*s*' and that they judge the theory by those general and philosophical considerations which, instead of concepts of a quantitative nature, now differentiate psychological schools. A distrust of the theory frequently arises because of the psychologist's lack of understanding of its mathematical foundations.) Spearman stated that another source of distrust had been the fact that the converse statement—that if hierarchy existed, '*g*' and '*s*' existed without any *o* [1]—had been unproved for so long. The definite and rigorous establishment of this proposition was Garnett's chief contribution.

Spearman then illustrated how transformations might disguise the presence of a general factor; he included an illustration fitting the data obtained from dice. The bulk of his paper is of no interest at present as it dealt with a criticism of the arbitrary arrangement of the dice in Thomson's first attempt at a hierarchy, and Thomson corrected most of these points in later papers. In concluding his review of

[1] Do not confuse *o* with Thomson's group factors or *d*'s.

the theory of '*g*' Spearman stated that, whatever the fate of the interpretative hypotheses, the main theory in abstract mathematical terms was definitely established and agreed to by both sides of the controversy.

In his paper (19) which reviewed the development of the controversy to 1920, Thomson noted that, in using Webb's data, Garnett had concluded that there were two large group factors which were nearly general. These factors were of the '*g*' and '*s*' type and had been called 'purpose,' *w*, and 'cleverness,' *c*. Thomson further argued that hierarchy was the natural order among correlation coefficients and that it was the absence of this phenomenon which necessitated explanation. He also discussed the relation of the problem of transfer of training to the theory of '*g*.' He stated that Spearman's theory would explain the absence of transfer very well if the theory would just add to itself the assumption that only the specific factors were influenced by training. According to Thomson's sampling theory, improvement in a function with practice is not due to improvement in the elemental abilities involved but is due to a selection of these abilities. The sample becomes more economical and unnecessary elements are discarded. In performing a different operation, the elemental abilities are just the same as if the training had not taken place. The new activity will be analogous to the first activity, and in the beginning it will use unnecessary elements. Transfer of training will then not take place except insofar as the second activity is a mere variant of the first. In this case a portion of the unnecessary elements have already been eliminated.

In 1921, Garnett reviewed (22) the 1921 edition of Brown and Thomson's 'Essentials of Mental Measurement.' The reviewer gave a compact statement of his former articles and of the cosine expression of correlation. He showed in detail the precise equations by which three variables could be factored into a '*g*' and three '*s*'s' in an infinity of ways; this is a result of the fact that four or more variables are required for the conditions of equiproportion, (1), under which a unique '*g*' and '*s*' are determinable. He attacked

Thomson's criterion (8) of a compulsory general factor in three variables on this ground, but he failed to admit that the criterion (8) is still the criterion of the necessary presence of general dice factors in any of the many ways in which correlated variables can be expressed by groupings of dice.

In another review (**21**) of the same book, Yule agreed that the intercolumnar correlation criterion is unsatisfactory and the formula for the correction for attenuation possibly misleading. He disagreed with Thomson's view that hierarchy is the natural order among correlation coefficients, and he stated that in psychological data even the uncorrected coefficients uniformly showed a marked tendency towards hierarchy but that in non-psychological data he had never observed hierarchy. He pointed out that the meaning of hierarchy would have to come from outside statistical studies, but using the evidence presented up to 1921 he concluded that "from the statistical standpoint, Dr. Spearman's explanation seems to me by far the simplest" (p. 105).

Privately, the reviewer learns that Thomson also prefers the '*g*' and '*s*' terminology as 'description' for his classes of teachers when they are measuring a child's ability or when they are predicting his success in school. Thomson prefers the view of the *d* elements, however, when he is thinking of causation or of the psychological influences which have given rise to the observed hierarchies. The *d* elements can account for any set of correlations between normally distributed variables in any fields where equiproportion is not found.

The next year, the theory was advanced by Spearman's proof (**23**) that if equiproportion existed, '*g*' and '*s*' alone existed. Garnett had proved this for normally distributed variables, but Spearman made the proof quite general. His proof began with the assumptions of Bravais and the use of Taylor's theorem and in the final stages worked into a complex determinantal expression. This the reviewer is incompetent to check, but he learns that Filon checked it minutely before transmitting it to the Royal Society.

In another article (**24**), Spearman pointed out in detail some ways, including that of dice, by which the general

factor, '*g*,' could be transformed and its presence concealed. He reaffirmed the position that Thomson's dice structures were not pure chance but were artificially manipulated to an extent that was tantamount to the introduction of a general factor.

By this time, the fact that equiproportion was explicable by either '*g*' and '*s*' factors or some arrangement of dice factors seemed apparent to everyone. The next step was the development of a criterion which would measure the divergency from perfect equiproportion more satisfactorily than the intercolumnar correlation.

In a further contribution (**25**) to the intercolumnar criterion, Spearman defended the assumption of lack of correlation between the quantities, ρ and ϵ, on the average of many tables. He also stated that Thomson had furnished no proof for his conclusion that the criterion cut off the larger values and produced an artificial average at unity, but that he had given only an illustration from a single table.

Then by an entirely different procedure in which he used Taylor's theorem through the third differentials, Spearman derived anew the formula for the intercolumnar correlation coefficient corrected for attenuation, and arrived at much the same result as he had previously obtained. He pointed out: (1) that Thomson's table was based on only 20 subjects; (2) that only five columns of it were even claimed to be usable; and (3) that in one of these five, the claim was based on an arithmetical error and that in another the intercolumnar correlation was calculated with a large error. He did say, however, "The corrective formula, though doubtless defective in the sense of only applying to a portion of the data, shows no sign of defectiveness in the sense of being invalid with regard to this portion" (p. 270).

The fundamental criterion of equiproportion, (1), can be written in the more convenient form,

$$r_{12}r_{34} - r_{13}r_{24} = O = F. \qquad (20)$$

Equation (20) is called the tetrad difference. In several

contributions (**27, 30, 37**), Holzinger and Spearman developed formulæ for the probable error of *F*. The formula which they gave as exact was long and cumbrous. For this reason, they gave several approximations to it. These formulæ seem not always to be entirely satisfactory in practice. They are for the probable error of a single tetrad. A table of the intercorrelations of a set of n variables, however, has ${}_3{}^nC_4$ tetrads. One may get an idea of the labor involved in calculating the probable errors of these individual tetrads by noticing that the table of correlation coefficients derived from a set of 14 variables would necessitate the calculation of 3,003 probable errors. In order to lighten the labor and to summarize the errors of sampling for the whole table, Holzinger and Spearman derived another formula (**30**) which they called the probable error of the distribution of tetrad differences. They state (**37**) that the square of this probable error "can easily be shown to equal the mean of the squared p.e.'s for all the tetrad differences taken separately" (p. xi). At the time of the writing of this review, their proof for the above statement has not yet been published. This formula, which is in terms of the average of the correlation coefficients and the sigma of these same coefficients about that average, has given, they state, good results in all the situations to which they have applied it, but they add that it is "on some theoretical points still awaiting elucidation" (p. xi).

Parallel with the attempts towards a better criterion of equiproportion, there were going on studies of the psychological properties of '*g*.' In his presidential address to the British Psychological Association (**32**), Spearman quoted data showing that some tests of the cross-out and tapping type involved more '*g*' at the feebleminded and early childhood levels than at normal and adult levels. This led him to propound a law of diminishing returns, those who possess larger amounts of '*g*' derive proportionately less from further increments of it. These conclusions are similar to those obtained by Hart and Spearman in their investigation (**4**) of dementia.

He stated that '*g*' was more important for complex

operations, and that '*s*' was more important for reproductive, for habitual, and for monotonous tasks. He argued that since the '*s*'s' were all independent, no such thing as a general sensory level should be expected. The practical consequence of this would be that in vocational guidance those '*s*'s' in which an individual excels should be discovered and trained. In this same article, he contrasted a set of correlation coefficients derived from psychological data with a set of coefficients derived from physical and sensory data. He gave the former as an example of a good fit to the criterion of the probable error of the tetrad differences for the whole table and the latter as an example of a poor fit to the criterion.

In the concluding three chapters of the 1925 edition of 'Essentials of Mental Measurement' (**29**), Thomson discussed the entire question of mental structure. Chapter IX outlined the theory of '*g*' and that of the intercolumnar correlation criterion. Chapter X included: (1) Thomson's sampling theory; (2) his evidence of hierarchies from dice structures; (3) his criticisms of the intercolumnar correlation criterion; and (4) his arguments for hierarchy's being the natural order of correlation coefficients. Chapter XI gave a review of the position of the controversy in 1924. Thomson stated that if hierarchy were present one had a choice between a general factor and group factors but that these group factors had to be peculiarly related. He agreed that if all the correlations were positive a general factor could always be postulated but that if the hierarchy were imperfect, group factors also had to be postulated. He noted that some tendencies towards hierarchies among the correlations obtained from economic and non-psychological fields would be found more often than Yule and others thought if these tendencies were quantitatively tested. He welcomed the tetrad difference criterion and pointed out that the vital need for progress was larger samples which would reduce the probable error of the divergence from perfect hierarchy.

In his next contribution (**33**), Thomson again asserted that the crucial difference between the theory of '*g*' and the sampling theory was that the former demanded the absence

of group factors. Here, Thomson committed the fallacy of not noticing that he and Spearman were using the word 'factor' in two different senses. Group dice factors in 'most probable' arrangements are entirely consistent with the theory of '*g*'; that which is demanded is the virtual absence of group factors after the dice factors have been transformed into the '*g*' and '*s*' factors. Therefore, the table of correlation coefficients which Thomson had derived by using only group dice factors should be expected to be equiproportional and to fit the tetrad difference criterion. These correlations in no way contradict the alternative analysis into a *unique* '*g*' and '*s*.' Thomson criticized the formulæ which Spearman and Holzinger offered (**27**, **30**) as approximations to the probable error of the tetrad difference. He repeated the calculation and distribution of the tetrads from Simpson's and Gates' tables. In the case of Simpson's data, he found more asymmetry and poorer correspondence to the theoretical expectation than had Spearman; in the case of Gates' data he found less asymmetry and closer correspondence than had Spearman. These differences arose because Thomson had used only half the available tetrad differences while Spearman had used them all. A symmetrical curve necessarily results from the latter treatment because for every tetrad there is another of equal magnitude but of opposite sign. Thomson also included a demonstration that equiproportion was to be expected by the conventional rules of probability.

In reply (**35**), Spearman pointed out that he was being misrepresented in that the absence of group factors is not demanded by the theory of '*g*.' (The reviewer has already pointed out that the word 'factor' was being used in two senses.) Spearman further claimed to have been the first to have discovered 'group factors' (p. 323) and to have led in the investigation of them. Both these factors and '*g*,' he grants, may be transformed into or from '*e*' elements, whether or not we can ever identify them with physiological processes. To the reviewer, it seems that these 'group factors' mar the perfection of equiproportion. If so, they are what we have called overlapping factors, *o*, and are residual after '*g*' has been eliminated.

In an unpublished paper, Dodd followed a suggestion of Spearman and developed a coefficient of equiproportion. This coefficient is a substitute for the laborious criterion of the average of the probable errors of the tetrad differences, and it summarizes the degree of equiproportion present, apart from its probability. If 'g' and 's' alone exist, the partials, $r_{ac \cdot g}$, from (2) will be zero, but if there is any o, this partial will measure its amount. The tetrad difference indicates the existence of an o, but neither are the variables in which it occurs designated nor is the amount of the overlap measured. In a set of n variables, there are 3^nC_4 tetrads but only $\frac{n^2 - n}{2}$ partials of the nature of (2). A finer analysis could be made by studying the distribution of these partials. In any case of complete analysis of the factors, these partials would already be calculated, so that but little extra labor would be involved in applying the coefficient of equiproportion. This is not the case with the tetrad difference criterion. The chief disadvantage of the partial r criterion is that it becomes proportionately inaccurate as a measure of the overlapping factors as the errors of sampling and observation become larger; though even under the latter situation, it indicates the overlapping factors as well as does the tetrad difference criterion. The partial correlation criterion [2] provides a test to determine whether imperfect equiproportion is due to sampling errors or to overlapping factors. As equiproportion approaches perfection, $r_{ac \cdot g}$ approaches zero, and its probable error approaches $\frac{1}{\sqrt{N}}$ as a limit. As N increases with larger samplings, the partial correlation coefficients will decrease proportionately to $\frac{1}{\sqrt{N}}$ if they are due to sampling errors, but they will become more stable, more constant, and larger than the probable errors if they are due to an o.

A contribution of great merit but of indirect bearing on

[2] This partial correlation was first suggested by Spearman who called it 'specific correlation.' His nomenclature seems to the reviewer to be unsuitable because factors that correlate are no longer specific to any one test.

the present review is the experimental work [3] recently published by Thorndike. The book contains much material for the theory of '*g*,' but no direct application has been made. In his review (**36**) of it, Spearman made several points which indicate that the various schools and views of intelligence are merging together in the direction of the single general factor analysis. After discussing Thorndike's definitions and 'dimensions' of intelligence, he pointed out that the criterion which had been used was essentially that of equiproportion. Thorndike decided to define intelligence for his purposes as functioning in four types of standardized verbal tests symbolized by the letters *CAVD*, and from the fact that these four tests correlated together as highly as attenuation errors would permit, he deduced that this functioning is a "unified, coherent fundamental fact in the world." [Thorndike, p. 484; (**36**), p. 368.] Thorndike further considered anything to be intellect *CAVD* that correlated with *CAVD* perfectly or as high as errors of measurement would permit. This criterion is a mathematically variant form of the tetrad difference, and it was this same measure which led Spearman to his publications of 1904 and 1907. The 'unified, coherent' functioning which Thorndike calls *CAVD* is but another name for '*g*.' The divergence of view between Spearman and Thorndike seems to be resolved in practice. These conclusions are of particular interest when compared with Spearman's and Thorndike's statements (**28**, **31**) of the problem (**26**).

So far, Spearman's two books have not been mentioned. The first dealt with the qualitative aspects of intelligence or cognition. It developed the three noegenetic principles of awareness, eduction of relations, and eduction of correlates, which with the anoegenetic principles of clearness, speed, and reproduction, are capable of describing the manner in which any cognition occurs. This book has been reviewed in several places, and as it does not develop the quantitative aspect of the theory of '*g*,' it will not be further considered here.

[3] Thorndike, E. L., et al., 'The Measurement of Intelligence,' New York, Teachers College, 1926, pp. xxvi + 616.

The second book (37) deals with the quantitative aspects of the theory of '*g*.' The first section of the book is a review and criticism of the doctrines of mental structure. This section shows the controversies and confusions that have arisen as a result of the lack of objective and quantitative definitions. In the second section, the tetrad difference is applied as the test between the different doctrines, and the author arrives at the theory of two factors.

He found '*g*' to be stable for a given individual and to be "involved invariably and exclusively in all operations of eductive nature, whatever might be the laws of relation or the sort of fundaments at issue" (p. 411). It seemed to be equally concerned with clearness and with speed which Spearman gives as the two dimensions of ability. It applies in a similar manner to extensity and intensity which are the two dimensions of span. Spearman characterizes span by the following law: "Every mind tends to keep its total simultaneous output constant in quantity, however varying in quality" (p. 259).

The discovery of a lag in mental activity gave rise to experiments by which a second general factor, *c*,[4] perseveration, was noted. This factor, *c*, is supposed to express the 'degree of inertia' (p. 306) of '*g*,' or the difficulty of changing from one task to another. He concludes that 'perseveration' and Garnett's 'cleverness' are "only the opposite aspects of one and the same thing" (p. 356). The individual's ability in memory or retentivity of dispositions was found to be particular to the kind of material retained, to be independent of '*g*,' but not to be a general factor.

The experiments of fluctuation led to a third single and universal factor, oscillation. This is "facility of recuperation after effortful expenditure" (p. 327).

The application of the tetrad difference criterion to correlations obtained from character ratings yielded a fourth

[4] It should be remembered that these letters are chosen as mere labels to signify approximately those quantities which possess objective and quantitative definitions. Verbal arguments about them are likely to be misleading, since the properties of the quantity may be different from the properties of the 'word' by which we name this quantity.

factor, *w*. If this factor is slightly shifted so as to be independent of '*g*' and '*s*,' it is Garnett's *z*. This factor is supposed to be something in the nature of will, purpose, persistence of motives, or self-control.

Of these four factors, the amount of '*g*' is the only one that figures appreciably in our ordinary tests of ability, intelligence, or school proficiency. The other three factors have been revealed only by specially designed experiments in fatigue, character rating, and fluctuation.

As an interpretation of these findings, Spearman presents a discussion favoring the hypothesis that '*g*' is some form of an energy and that the '*s*' is more likely the particular neurone pattern involved in any measured ability. Under this hypothesis, '*g*' is the amount of the energy, *c* is its inertia, oscillation is its variation—probably a manifestation of fatigue—and *w* is that which 'has a dominating influence upon all exercise or even estimation of cognitive ability' (p. 413).

At the time of writing this review, Thomson had but recently published a paper (**34**) on the possible linkages of four correlated variables and the resulting tetrads. His first example contained 108 elements, and his first variable contained a sample of 90 of these elements, his second variable a sample of 72, his third a sample of 54, and his fourth a sample of 36. He allowed the variables to have overlapping elements in the most probable manner. Thus in variables one and two, there would be $\frac{90}{108} \times \frac{72}{108} \times 108$ or 60 elements which would be common. He calculated the theoretical correlations, formula (7), and found that the tetrads were zero.

He next took a universe of 6 elements and distributed it in such a manner that the first variable had 5 elements, the second 4, the third 3, and the fourth 2. This system gave him 78 different patterns of linkage but only 76 different sextets of correlation coefficients. The total frequency of these 76 different sextets was 450, if it be so arranged that 1 is the frequency of the least probable sextet. The tetrads

from these correlations gave a distribution of which the average was zero and the variance .04. He then took a universe of 12 elements and doubled the number of elements in each variable. This system was handled as the two previous ones, and the tetrads gave a distribution, the average of which was again zero but the variance of which was only .018 or about half of the variance in the universe with half as many elements. He concluded that "correlation coefficients tend towards zero tetrad-differences without any need of the assumption that only a general factor and specific factors exist, to the exclusion of group factors" (p. 76). He qualified this conclusion with the condition that the underlying connections have the all-or-none property, but he believed that the conclusion would stand without this qualifying condition.

However, it is not the average but the dispersion of the observed tetrads to which one refers when one speaks of the 'tetrad difference criterion,' that is, one wishes to know if the observed tetrads have a larger sigma than can be reasonably attributed to variations due to sampling. In Thomson's experiments, which were just cited, the correlations had been calculated from a formula in which the N was very large so that he was dealing with the whole population or true correlations only. Thus, in the case of every correlation coefficient, the probable error due to sampling was zero, and the probable error of the tetrad differences was zero. The tetrad difference criterion would accept neither of his last two examples as being equiproportional because his observed variances were .04 and .018. In this case, then, the criterion would show that the variables could not be divided exclusively into 'g' and 's.'

Thomson could have taken a single sextet and argued that just as a sampling variation is introduced by using a population less than the whole population, so a sampling variation would be introduced by using a number of elements smaller than the whole number of elements. The formula [5] for the probable error due to this sort of sampling has since

[5] Mackie, J., 'The Tetrad Difference Criterion,' *Nature*, 1927, 120, 298.

been determined. Thomson, however, worked out not a few instances of linkage but all of them in the proper relative frequency. This is equivalent to using a very great number of elements and not a sample of them. Hence even by this type of variation, there should be no distribution of the tetrads around zero if the resulting correlations were in equiproportion.

Thomson's article, then, with its first example illustrated Garnett's proof of the statement that perfect equiproportion will be obtained from variables composed of dice-like factors distributed in a most probable manner and that these variables can be alternately expressed in terms of '*g*' and '*s*,' or *e*'s. The '*g*' and '*s*' exist and exist exclusively whether or not one cares to make the transformation. His other two examples illustrate the fact that all tables of correlation coefficients tend to give tetrads the average of which is zero, but that only those correlations are equiproportional which give tetrads the dispersion of which may be attributed to sampling. His article is of particular importance and usefulness as an illustration of how to develop and use a situation which is free from the complications of sampling.

Spearman and Dodd both have unpublished papers in which the sampling theory of intelligence is further investigated. Spearman had assumed that the tests sampled elements in such a manner that the most probable number of elements sampled by any one test was equal to the most probable number sampled by any other test. This hypothesis led to a set of equal correlation coefficients; in this set equiproportion still exists, but the intercolumnar correlation would be zero. Thus Spearman drew the conclusion that the number of elements could not be very large.

In investigating several possible sampling conditions, Dodd followed the general leadership of Spearman but allowed different tests and different degrees of freedom for sampling the very numerous elements. The psychological interpretation of these hypotheses would be that some tests depend on more elements than others. Dodd's assumptions may be stated as follows: (1) the observed variables can be expressed

as additive functions of numerous small elements—sigma units are used throughout; (2) the coefficients of the elements are 1 or 0; (3) the proportion of elements sampled varies from test to test, but for any one test this proportion is constant for all individuals. If these assumptions are to stand, they must explain the facts that the observed variables tend to be distributed normally, that the obtained correlations tend to be medium, positive, equiproportional, and hierarchical. A mechanism is shown in detail for artificially synthesizing the observed variables under the assumptions. The consequences of the assumptions are followed through in detail, and formulæ are developed for the relations between the number of the elements, the sigma of the elements, the sigma of the observed variables, and the sigma of '*g*.' In harmony with Spearman's conclusions, it was found that the number of elements could not be very large, but that the assumptions satisfactorily explained all the facts if the number of elements was of the order of hundreds or even smaller. If the number of elements is very large, there will be no dispersion of '*g*,' but this objection was met in the smaller number of elements by actual experiments, and by the formulæ which show that the sigma of '*g*' can be made to correspond to the observed data by appropriate distributions of the elements. These assumptions are then compared with Thomson's dice structures and with his hypothesis that hierarchy is due to a correlation of sampling errors. There is then made the addition of interchangeable elements which had been suggested verbally by Thomson but which had never previously been included in the mathematical treatment. This suggestion is that chance determines the elements which any one individual uses in any test. Thus, two individuals may perform the same task, may get the same scores, but use different elements.

There are then worked out the probable errors of the correlation coefficients for the sampling of the population and for the sampling of the elements by each individual. It was found that the variation due to sampling in no way conflicted with equiproportion or other observed quantitative facts. Interpretative hypotheses are then reviewed.

The formula for 'g' is

$$g = e_1 + e_2 + \cdots + e_n = \Sigma(e) \qquad (21)$$

and the sigma of 'g' is $\sqrt{n}\,\sigma_e$ so that equation (21) reduces to (18) if one makes the sigma of 'g' equal to the sigma of 'e.' Thus 'g' is the sum of the elements as the foot is the sum of 12 inches; whether 'foot' or 'inch' forms the 'fact' depends upon one's purpose.

At this point, the theory of factors changes from the particular case of equiproportion to the more general case of criteria for factoring. There are being developed the general equations, together with their probable errors, which give the conditions under which the observed variables may be expressed in terms of general, overlapping, or specific factors. These factors are for the purpose of analyses, trials of hypotheses, and for concepts for the quantitative systematization of masses of observed data. They in no way reveal causation; causal factors must be inferred on other grounds, though even here a factorial theory is helpful if it is properly manipulated.

One step in this direction, was Dodd's generalization (38) of Garnett's criteria for factoring three variables into two general factors and four variables into three general factors without residual group or specific factors. The generalized criterion for n variables to be factorable into $n - m$ factors is that every multiple correlation ratio or coefficient of the $(n - m)$th order shall equal unity. In the case of rectilinear regression, the more convenient criterion is that every discriminant of the $(n - m)$th order shall equal zero. This discriminant is the determinant formed from the table of intercorrelations with 1 as each of the self-correlations in the cells of the main diagonal. The value of this determinant is always less than unity. The discriminantal criterion is to be preferred where applicable, as there are $n - m + 1$ times as many multiple correlation criterions of which the approaches to perfection have to be judged in any given problem.

With the aid of Pearson, the probable error of the nth

order discriminant, Δ, was found to be

$$\text{P.E.}_{\Delta} = (.6745)2\Delta\sqrt{\frac{r_{12}+r_{13}+\cdots+r_{23}+\cdots+r_{(n-1)n}}{N}}. \quad (22)$$

This revealed that any observed value of Δ in excess of zero will always be many times its probable error because the radical is always a small fraction unless the number of variables approaches the number of subjects, and this never happens in satisfactory experiments. Thus, any failure to satisfy the criterion cannot be attributed to sampling errors, as these will always play a relatively negligible rôle, but the errors of observation will always remain.

The criterion that the discriminant shall vanish

$$\Delta = 0 \quad (23)$$

or that the multiple correlation ratio shall be perfect

$$\eta_{1\cdot 23\cdots(n-m-1)} = 1.00 \quad (24)$$

is also the crucial test for the complete absence of specific factors of any type in any of the variables. The perfect prediction of any variable is impossible if that variable contains any factors not common to the other variables.

It is always possible to factor n variables into n other variables. It is most satisfactory, however, to choose the n new ones so that they are independent of each other. The following advantages ensue: (*a*) the geometrical representation of the independent factors is that of an orthogonal coördinate system which enables the connecting equations (10) to be in terms of direction cosines; (*b*) the squares of the correlation coefficients may be considered arithmetically as percentages; (*c*) partial correlation then accomplishes the clean elimination of a factor; (*d*) the definitions of the new factors fulfill the canon of mutual exclusion; and (*e*) causal interpretations and hypotheses are aided.

It is also possible to factor n variables into $n + m$ factors in an almost unlimited number of ways unless further conditions are imposed. There have been devised rules and

schemes by which variables with any given intercorrelations may be constructed from groupings of dice-like elements. Thompson has done this (**15**) for three variables, and Thomson has done it (**39**) for four.

In another unpublished paper, Dodd attempted to generalize the conditions for a compulsory general factor in any number of variables. He worked with the simplest possible type of transformation and showed how many orders of factors could exist for any number of variables. The next step was the development of the critical equations analogous to (8). The limit of this type of analysis is sharply defined at the point where the number of unknown and possible factors exceed the number of known correlation coefficients from which they are to be determined. Theoretically, it is always possible to work out the criterion for the compulsory presence of a general factor, but practically, the equations become very complex.

In 1919, Thompson published a study (**15**) which dealt with interference factors. An interference factor is a factor which is common to at least two variables and which is positively correlated with one variable and negatively correlated with another. This type of factor allows a vastly larger assortment of mechanisms producing correlation. It is even possible that two variables have several common elements and yet show no correlation if these elements are of opposite sign and of equal magnitude. This type of mechanism needs more investigation, for it might explain many of the anomalies arising from experiments on practice and transfer of training. If training were to act differentially on the positive and interference factors, the intercorrelations would show the most tangled and superficially contradictory results.

Summary

All the original contributions to Spearman's theory of two factors, '*g*' and '*s*,' and to Thomson's sampling theory have been reviewed. After consultation with both of them, a more exact terminology has been introduced in order to reduce ambiguities; hierarchy has been more precisely defined

by equiproportion. All the chief parties to the controversy recognize the fact that to the extent that n correlated variables are in equiproportion they can always be factored into a uniquely determined general factor, 'g,' and n specific factors, 's.' They also agree that in such cases of equiproportion, the variables can be factored into numerous independent elements, e, which are distributed according to some system of probability. Each system involves further assumptions of which the legitimacy may be open to question. With these qualifications, the two systems are interchangeable; both exist simultaneously. In interpreting these systems into psychological and physiological terms, conflicts may occur, and one system may preclude the other. For the purpose of measurement and prediction, the concepts of 'g' and 's' are the more useful. There is no evidence bearing on their agreement with physiological concepts. The system of e elements may prove very useful in situations where equiproportion does not hold. In Spearman's two books, there is a large body of evidence for the interpretation of equiproportion by means of 'g.' The theory is being extended from the particular case of equiproportion to the more general theory of the factors underlying correlated variables.

REFERENCES (CONTINUED FROM PART I)

18. SPEARMAN, C., 'Manifold Sub-Theories of "The Two Factors,"' PSYCHOL. REV., 1920, 27, 159–172.
19. THOMSON, G. H., 'General versus Group Factors in Mental Activities,' PSYCHOL. REV., 1920, 27, 173–190.
20. THOMSON, GODFREY H., 'The General Factor Fallacy in Psychology,' *Brit. J. Psychol.*, 1920, 10, 319–326.
21. YULE, G. UDNY, 'Critical Notice of "The Essentials of Mental Measurement,"' *Brit. J. Psychol.*, 1921, 12, 100–107.
22. GARNETT, J. C. MAXWELL, 'Review of "The Essentials of Mental Measurement" by William Brown and Godfrey H. Thomson,' *Brain*, 1921, 44, 332–337.
23. SPEARMAN, C., 'Correlation between Arrays in a Table of Correlations,' *Proc. Roy. Soc. Lond.*, 1922, 101A, 94–100.
24. SPEARMAN, C., 'Recent Contributions to the Theory of "Two Factors,"' *Brit. J. Psychol.*, 1922, 13, 26–30.
25. SPEARMAN, C., 'Further Note on the "Theory of Two Factors,"' *Brit. J. Psychol.*, 1923, 13, 266–270.
26. SPEARMAN, C., 'The Nature of "Intelligence" and the Principles of Cognition,' London: Macmillan, 1923. Pp. viii + 358.

27. SPEARMAN, C., AND HOLZINGER, K., 'The Sampling Error in the Theory of Two Factors,' *Brit. J. Psychol.*, 1924, **15**, 17–19.
28. SPEARMAN, C., AND THORNDIKE, E. L., 'A Communication and a Reply,' *J. Educ. Psychol.*, 1924, **15**, 393–395.
29. BROWN, WILLIAM, AND THOMSON, GODFREY H., 'The Essentials of Mental Measurement,' London: Cambridge Univ. Press, 1925. Pp. 224.
30. SPEARMAN, C., AND HOLZINGER, K., 'Note on the Sampling Error of Tetrad Differences,' *Brit. J. Psychol.*, 1925, **16**, 86–88.
31. SPEARMAN, C., 'Agreement on Coöperation,' *J. Educ. Psychol.*, 1925, **16**, 423–425.
32. SPEARMAN, C., 'Some Issues in the Theory of "*g*,"' *Proc. Brit. Assoc. Adv. Science*, 1925, Sect. J.
33. THOMSON, GODFREY H., 'The Tetrad-difference Criterion,' *Brit. J. Psychol.*, 1927, **17**, 235–255.
34. THOMSON, GODFREY H., 'A Worked-out Example of the Possible Linkages of Four Correlated Variables on the Sampling Theory,' *Brit. J. Psychol.*, 1927, **18**, 68–76.
35. SPEARMAN, C., 'Material versus Abstract Factors in Correlation,' *Brit. J. Psychol.*, 1927, **17**, 322–326.
36. SPEARMAN, C., 'Critical Notice of "The Measurement of Intelligence" by Professor E. L. Thorndike and Staff,' *Brit. J. Psychol.*, 1927, **17**, 365–369.
37. SPEARMAN, C., 'The Abilities of Man,' London: Macmillan, 1927. Pp. vi + 415 + xxxiii.
38. DODD, STUART C., 'On Criteria for Factorizing Correlated Variables,' *Biometrika*, 1927, **19**, 45–52.
39. THOMSON, GODFREY H., 'On the Formation of Structure Diagrams between Four Correlated Variables,' *J. Educ. Psychol.*, 1927, **18**, 145–158.

[MS. received November 19, 1927]

A SUGGESTIVE REVIEW OF *GESTALT* PSYCHOLOGY

BY H. H. HSIAO
Berkeley, California

I. Introduction

A survey of the English literature on the psychology of *Gestalt* easily makes one realize the fact that there has, as yet, been no clear, simple, systematic and unprejudiced presentation of the subject. Although, in addition to the two translations, Köhler's 'Mentality of Apes' and Koffka's 'Growth of the Mind,' quite a number of articles have been published in the various journals, most of them are in nature critical rather than presentative. Among the rest, the only articles that contain a rather comprehensive view of the new psychology are those by H. Helson. Unfortunately, in the last-named, the presentation is such that it is apt to leave upon the mind of readers only a chaotic conception of the subject. However, Helson himself has confessed that the nature of the material makes a systematic treatment of it difficult.

These facts have led the writer to realize the desirability of an attempt to re-present the subject in a different manner, however inadequate it may be. He will also take the opportunity to consider the most important attacks that have been made upon this psychology and offer a few suggestions of his own.

II. The Concept of *Gestalt*

It goes without saying that the concept of *Gestalt* (configuration) is the key to the understanding of *Gestalt*-psychology; but, unfortunately, so far there has been no attempt to clarify this concept to the extent that is possible. It is this failure that has led to many a misunderstanding. Now, in order that such may be removed or prevented, it is

necessary, I believe, to present the various aspects of the concept instead of giving it a mere definition, which is apt to be abstract and ambiguous. For this purpose, the following aspects of *Gestalt* are chosen partly owing to their importance to the understanding of this concept and partly owing to their applicability to most, if not all, forms of experience.

The Concept of Phenomenal Pattern.—From the results of their experiments, the *Gestalt*-psychologists have concluded that what is perceived is not always the same as the stimulus. The former is referred to as phenomenal pattern. This distinction will be made clear by one of Wertheimer's experiments. Two black lines were exposed in succession on a white background, the one vertical, the other horizontal, in such a fashion, that, if they both were seen simultaneously, they would form a right angle. With slow rotation of the stroboscope, first the one line appears at rest, and then the other. Quick rotation causes both lines to appear stationary at the same time. If the speed is adjusted between these two rates, the one appears to move over to the other, and often returns to its original position. This is Wertheimer's optimal movement, and is known as 'phi-phenomenon.' Here the 'phi-phenomenon' is a phenomenal pattern, while the two lines are stimuli.

The Concept of Figure and Ground.—A *Gestalt* always consists of two parts, figure and ground. They represent two fields, the one of which predominates, while the other sets it off. The figure has a definite structure, whereas the ground forms the background of the figure. The former is always more resistant to change than the latter. An object is not perceived until the field has developed into a figure.

Fig. 1

Phenomenologically, figure and ground possess different properties. A given optical pattern becomes two entirely different things when perceived now as ground and now as

figure. For instance, in looking at the accompanying picture, if we make the white parts figure, then the black parts become ground, and vice versa. This shift from one to the other is accompanied by changes in the phenomenal properties of both white and black parts, for they are not the same when seen as figure and when seen as ground.

The Concept of Wholeness.—The *Gestalt* is a whole in the sense that it is not a mere summation of parts; for, in it, there is something over and above the parts summated. For instance, a square is not merely a summation of four straight lines; it means more than such a summation. Therefore a *Gestalt* is governed by internal laws as opposed to the summative, contingent, spatio-temporal contiguities of elements.

The Concept of Transposability.—*Gestalten* are transposable in so far as they possess similar essentials. For instance, Köhler trained a number of apes to pick their food from the darker of two gray containers. When a third container, darker than that to which the animals had responded, was introduced, the animals picked the darker of the last two present. Here it is the similar characteristic "darker than" that determines the transposability of the two *Gestalten.*

The Law of Prägnanz.—According to the Law of Prägnanz (precision), a *Gestalt* tends to become as sharply defined as possible. For example, Köhler, working with sound intensities, found that the second of the two objectively equal stimuli was heard, not by itself alone, but as a rise or fall from the first, that is, as a step from the first. Such a stepwise perception tends to appear wherever conditions are at all favorable for it. The reason is that a step is more impressive than an undifferentiated field. This shows how the Law of Prägnanz works.

The Principle of Closure.—According to the Principle of Closure, in every process, which issues at all in an end situation, the mode of distribution of energy shifts in the direction of a minimum of *Gestalt* energy. Fig. 2 appears at once as an open triangle, although it does not have three angles. Speaking in terms of this principle, this figure,

Fig. 2

although not closed, yet indicates, with a relatively high degree of certainty, the direction in which closure is to be effected.

The Principle of Intimacy (Innigkeit), Organization and Formation.—The Principle of Intimacy refers to the fact that a *Gestalt* is not an aggregate from which parts may be subtracted or to which parts may be added without affecting its other parts. On the contrary, in 'an intimate multiplicity,' the independence of parts is lost, for none of them can undergo a change without affecting the rest. For instance, if a certain number of men form a circle, it is impossible to admit an outsider as a member without affecting the positions of the other members as well as the whole.

Organization refers to the state of affairs in the *Gestalt.* Through organization a state of relief is produced, in which superior members take the lead, whereas the others stratify themselves. This state of affairs since Plato has been compared to the society, which resembles a *Gestalt*, not only in a few particulars, but in the structure as a whole.

According to the Principle of Formation, the *Gestalt* is plastic; and, through the process of formation, it 'comes more or less to itself.' A complete formation determines the *Gestalt*, and makes it decisive. For instance, a triangle is the *Gestalt* that fulfills a definite geometrical law of formation; but some triangles are triangles in a higher degree than others, *i.e.*, they are characteristic or normal triangles. This is true of all *Gestalten*, and therefore it is always logical to speak of a thing as more or less of a *Gestalt.*

It should be born in mind that the concepts, laws and principles that are here summarized do not all represent absolutely separate factors or aspects of a *Gestalt.* For instance, the Law of Prägnanz and the Principle of Closure are almost identical; and among such concepts as wholeness, intimacy and phenomenal pattern, there is undoubtedly much

overlapping. Being thus placed side by side, I believe, the amount of their overlapping may be readily determined. A bird's-eye view of the comprehensive concept of *Gestalt* is also thereby made possible.

III. Psychological Processes Viewed in the Light of *Gestalt-Theorie*

Before viewing the different psychological processes in the light of the *Gestalt-theorie*, we have to make clear the meaning of the two terms 'temporal' and 'non-temporal' *Gestalten*. The former arise from the succession of the factors that constitute it, the latter from the simultaneity of their appearance. A song is an instance of the first sort, while a picture is an instance of the second. In fact, certain psychological processes are concerned with temporal *Gestalten*, others with non-temporal ones and still others with both of them. One will do well to keep this fact in mind in considering the *Gestalten* that are involved in the various psychological processes.

Perception.—Regarding the nature of perception Wertheimer says, "Perception does not arise from the sum of individual stimuli on the one side and a corresponding sum of sensations on the other under secondary summative addition of further factors, but of the stimulus-constellation on the one side and the psychologically actual event in the *Gestalt* form on the other. Besides the factors of the stimulus-constellation, the subjective factors which are governed by laws are also decisive. It is these facts that represent characteristic totalities in an essential way."[1] This is a common notion of perception within the leading group of *Gestalt*-psychologists.

In the discussion of the concept of figure and ground, I have already mentioned the fact that a thing will not be perceived until it has developed into a figure. As to the perception of its parts, the possibility of their being perceived depends upon their position within the whole, whether or not they are natural parts and will lend themselves to membership

[1] *Psychol. Forsch.*, 1921–22, 1, 54.

in a new *Gestalt* and upon a number of other factors such as persistence, impressiveness, etc.

So far, *Gestalt*-psychologists have studied only four senses, namely, visual, auditory, tactile and olfactory. The richest fruits have been reaped in the field of visual perception; but, as even to summarize these results would require another paper, here I will call attention to no more than the fact that the whole *Gestalt*-movement has originated in such studies, especially in Wertheimer's study of the apparent movement, which has already been briefly touched upon in this paper. It is the results of this study that have led him to the conclusion that the perception is not analyzable into single sensational units, and explainable by means of them. His main conclusions have been supported by subsequent investigations.

As to auditory perceptions, they, as *Gestalten*, fall into two groups: temporal and non-temporal. A single tone or chord perceived instantly represents a non-temporal *Gestalt*, whereas themes and melodies that are heard in a choral or symphony, as they require time for their completion, are regarded as temporal Gestalten.

The concept of figure and ground has also been applied to auditory perceptions. For instance, in a musical performance, there are two kinds of intervals: the inter- and the intra-metric. The intra-metric intervals belong to the rhythmic group itself, *i.e.*, to the figure, while the inter-metric ones belong to the ground. This group may be stillness or a mixture of noises in the surroundings.

Repeated experiments have shown that, upon the presentation of two short, sharply-bounded, 'point-like' auditory stimuli 'a' and 'b,' a movement-phenomenon may be produced like Wertheimer's pure 'phi.' This movement is heard from 'a' to 'b.' It is not 'a' that moves; it is a movement without a thing. The time between the two sounds is not filled in with any sound sensation; but, during this favorable pause, the movement appears.

The study of olfactory perceptions in the light of the *Gestalt-theorie* had been neglected until 1926 when it was taken

up by J. Herrmann.[2] He has found that the smell experience is imbedded in heterogeneous complexes. These comprehensive totalities of smell experience are richly stratified. As most of them are sharply characterized by complex qualities, they are distinguishable and recognizable with exactness.

Herrmann has also pointed out the fact that, in smelling, it is not a small number of brain nerves, but the whole man that is involved. The individual character of a smell rests as much upon the attitude of the person, as that of a song or speech.

In regard to the tactile-*Gestalten*, only one thing is worth mentioning for the present, *i.e.*, the phenomenon of tactile 'phi' or apparent movement on the skin. This phenomenon has been studied by Benussi and Scholz, and striking similarities have been established between touch and vision.

Attention.—*Gestalt*-psychologists deny attention the following three functions, namely, as synthesizer, for this is not justifiable by observation; as threshold-determiner, for it is the *Gestalt* that determines the threshold; and as a term to denote clearness, for changes of clearness also involve changes of the *Gestalt*. According to the new theory, what will be attended to depends upon the *Gestalt*, and is explainable from such factors as position within a structure, whether a field is figure or ground, and how the *Gestalt* conforms to laws of Prägnanz and the like.

Recognition.—Recognition owes its possibility to two factors. One of these is the transposability of the *Gestalt*. For instance, a triangle may be recognized wherever its geometrical qualities are present regardless of its size or color. The other fact is that all our experiences conform to *Gestalt*-laws, and therefore the conditions for recognition are present therein.

It goes without saying that the form of a *Gestalt* is an important factor in recognition, but its properties are just as important. Experiments have shown that when the

[2] Herrmann, J., 'Gesamterlebnisse bei Gerüchen,' *Neue Psychol. Stud.*, 1926, 1, 473–506.

formal properties of different sense modalities do not coincide, identical stimuli appear as different objects, and may pass unnoticed. Furthermore, if the field has formerly been seen as figure and is later to appear as ground, it will not be recognized as the same, for figure and ground possess entirely different phenomenal properties.

Memory.—According to Wertheimer, memory is concerned above all with the properties of totalities and connections of structures. The essence of memory does not consist in the joining of sensations, results or wholes, if these wholes are merely summative partial wholes (Stückganze).

As to the general law of memory, it is stated by *Gestalt*-psychologists in the following way: if a new *Gestalt* arises under fixed objective conditions, it will be somehow preserved in the organism, and will arise more readily than before when the objective conditions are repeated. The *Gestalt*, however, will tend to reappear even when the conditions that are to call it forth are incomplete.

To this law one more fact at least has to be added to complete the *Gestalt* conception of memory. This is reported by F. Wulf as follows: Just as not every *Gestalt* will be perceived, so not everything that is perceived will be retained in memory. That which remains in memory, the physiological 'engram,' is accordingly not to be looked upon as an unchangeable impression, which, only in the course of time, would gradually fade away like carvings on paving-stones. This engram undergoes many more changes on the basis of *Gestalt* laws. In place of the originally perceived *Gestalten*, changes occur in certain respects in the course of time, and these changes concern the *Gestalten* as a whole.

There are in general two tendencies of changes, namely, accentuating and levelling. They seem to be opposed to each other, but they have a common feature in the fact that these changes are both in the direction of a better *Gestalt*. However much sharpness and inner organization may change with the lapse of time, the tendency towards a good *Gestalt* is not thereby affected. The tendency of levelling by smoothing over the asymmetry as in the case of the pictures to be

memorized in Wulf's experiment certainly leads to a better *Gestalt*, and so does the accentuating tendency.

Motor Acts.—Every motor act appears as a figure upon the general background of the total organism. Even such isolated acts as the raising of an arm and the singing of a note involve the tonic level of the whole muscular and nervous system. Moreover, every act leaves a more or less permanent effect upon the organism in the form of a *Gestalt*-disposition, and is modified by the past, present and future needs of the organism in the economy of the whole.

Thought Processes.—Thought processes, *i.e.*, processes involving the original solution of a problem, are to be distinguished from mere memory processes with which they seemed to be irrevocably united. They are also to be distinguished from processes of piecemeal generalization, piecemeal subtractive abstraction, combination, etc. They show themselves essentially as 'concrete-characteristic,' definitely conditioned *Gestalt*-processes, which find their analogies in processes of perception on the one side and in those of will or feeling on the other.

Speaking in terms of the Principle of Closure, we might say that the problem that demands thought arouses a *Gestalt*-process, which is incomplete and calls for closure. Once the *Gestalt* demanded by the problem is understood, the inner conception appears, and then the *Gestalt* is completed. As to creative thinking, it necessitates not only the perception of a given *Gestalt*, but also the destruction of old ones in order that the new *Gestalt* may become dominant.

Instinct.—Instinct is a temporal *Gestalt*, for it has an initial, a transitional, and an end stage. So long as the instinctive activity is incomplete, every new situation created by the animal is to it a transitional situation. It is the goal striven after that constitutes the end stage of the instinct, so the activity of the animal will not cease until the goal is reached.

Another noteworthy characteristic about the instinct is that every member of the instinctive activity is determined not only by its position with reference to what has gone

before, but with reference to all the members of the complete act, and, particularly, the last. In these respects, the instinctive activity resembles such phenomena as figure, rhythm and melody.

Intelligence.—There are three criteria of intelligence which are emphasized by *Gestalt*-psychologists. One of these is the manipulation of things with reference to the *Gestalt* in which they stand to one another. For instance, in a special series of experiments, Köhler employed four tubes of different diameters so that No. 1 fitted into No. 2, No. 2 into No. 3, and No. 3 into No. 4. Two of these tubes chosen in a serial order as given were laid horizontally parallel before the animal, sometimes the thinner and sometimes the thicker being near at hand. In the majority of cases, the animal picked up No. 2 with the left hand, or with the right hand, according as it was to be combined with No. 1 or No. 3. In other words, the animal handled the objects with reference to the *Gestalt* in which they stood to one another.

Another criterion of intelligence is the relatedness of the different parts to the whole activity. This may be illustrated by the following experiment. A chimpanzee was confined inside the bars, and a short stick was left with him. Outside the bars, a long stick was placed at the distance that could be reached only with the short stick, and some fruit was placed at the distance that could be reached only with the long stick. The chimpanzee proved able to secure the long stick with the short one, and to pull in the fruit with the former. These two acts—that of getting the long stick with the short one, and that of getting the fruit with the long stick—are essentially related with reference to the whole activity. It is this relatedness of these acts that is used as another criterion of intelligence.

The ability to reorganize the field to satisfy a need is also an index of intelligence. The act on the part of Köhler's chimpanzee to move a box under some hanging fruit so that the latter might be reached by standing on the box is an instance in point. As we know, some of Köhler's animals succeeded in putting four objects one upon another; some

scarcely got beyond constructions composed of two boxes; and others never built. These differences surely indicate differences in intelligence.

However, these criteria of intelligence are not meant to be independent of one another. Very often they refer to the various aspects of one and the same intelligent performance, although these aspects may differ in their relative importance in different acts.

Habit.—In connection with Habit, two things may be mentioned, *i.e.*, the difficulty of its formation and the function of repetition. According to the *Gestalt-theorie*, it is owing to the fact that the conditions of the external surroundings or of the internal organization of the animal exclude the possibility of immediately apprehending the *Gestalt*, that a long time is required for the process of habituation. Under such conditions, the act has to be repeated again and again in order to bring out the *Gestalt*. Instead of serving to strengthen bonds, the chief function of repetition is to prepare the ground for the construction of an appropriate figure.

This assumption agrees well with the known facts. For instance, in a purely habitual achievement, like that of mechanically learning a series of nonsense syllables, a collective apprehension is requisite, in which several members are united into a whole. This construction of a unity often occurs in the form of rhythmic groups, and, in general the material, in order to be learned, has to form some sort of figure. This essential point has been confirmed by the findings of A. Kühn and K. Lewin.

Learning.—In opposition to the associationist conception of learning as a linking of various psychic pictures, *Gestalt*-psychologists regard learning as performance of a higher degree, which comprises a group of practices leading to a definite result and creates an ability to do what was once beyond one's power. To this conception of learning also belongs the preëxistence of an aim to develop this ability through practice.

In his 'Growth of the Mind,' Koffka discusses learning under the following four headings: sensory, motor, sensori-motor and ideational.

Under sensory learning the learning of colors and forms is included. The learning of color names is said to depend upon the possibility of arousing an accurate color *Gestalt*. Color difference is apprehended only when there arises a *Gestalt* of two colors in which they appear as in a given *Gestalt*. The development of color-perception is therefore the gradual construction of new color-figures. As to the learning of forms, it has been shown that the figures of early infantile perception may be readily aroused, though they are very crude and possess little in the way of internal structure. However, the internal structure becomes constantly more definite as the child develops.

In motor learning, a formation after the manner of a *Gestalt* takes place, which, although composed of independent parts, is yet an articulate whole. Therefore, the improvement in a performance is the construction of a better and more complete *Gestalt*. Such an improvement is not an intelligent activity, for to know beforehand how the act is to be performed is of no avail. This fact suggests that a *Gestalt* of this kind does not originate as one that is due to intelligence; its neural basis must lie in other centers. But, in beginning to learn, the phenomenon of perception and a fixed purpose are essential, as the formation of a *Gestalt* is influenced by these components. Here practice has a two-fold function: to fixate the behavior and to create conditions favorable to the arousal of the new *Gestalt*.

As to the sensori-motor learning, Koffka cites, as an example, the maxim that "A burnt child shuns the fire." He says that the chief function of pain in this experience is to arouse attention, and thus furnish favorable conditions for the construction of a new *Gestalt*. The instances cited in connection with the discussion of intelligence are also illustrations of the sensori-motor learning.

It is through ideational learning, Koffka says, that man frees himself from the perception of things present to his senses, and thus attains his mastery over the world. Under this heading he places the learning of words and numbers. Words are said to enter the thing-*Gestalt* just as sticks enter

the animal's situation of 'desire to get fruit.' As to the learning of numbers, the formation of natural groups is, according to Wertheimer, prior to counting; and serial counting and group construction remain different processes even after the child has learned to count properly.

So far, the different psychological processes have been considered as such; but attention should now be called to a theory which is fundamental to all of them. According to *Gestalt*-psychologists, perception is an essential factor in all psychological processes. They depend upon and vary with it. But perception in turn rests upon the state of the whole organism, which itself seems to be regarded always as a basal *Gestalt*. This notion is apparent in Koffka's assertion that the physiological process is structural, and Köhler is also of exactly the same opinion. Under conditions of stimulation, this basal *Gestalt*, according to its laws, brings forth definite forms of perception or performance. This is a point often overlooked by previous critics of *Gestalt*-psychology.

IV. Criticisms and Suggestions

With the above facts in mind, we are now in a position to consider important criticisms of *Gestalt*-psychology. By doing so, some of the points that have failed to be touched upon in the preceding paragraphs may be, to a certain extent, cleared up.

The most fundamental defect attributed to the new psychology is its neglecting the study of the properties and laws of the experiencing individual. But the very emphasis that is always placed by this school of psychology upon such factors as need, goal, attitude, phenomenal pattern and the like is itself an argument strong enough to refute this accusation. Moreover, as *Gestalten* are the direct experiences and responses of the experiencing individual, it is natural that, through the study of these, the exact knowledge of his properties and laws may be obtained.

Another criticism concerns the objection of *Gestalt*-psychologists to the method of analysis. In this connection, it should be clearly understood that the kind of analysis to

which they object is the analysis of experience into elements that are not directly experienced in every-day life. As such an analysis is artificial and useless, the substitution for it of functional analysis is not uncalled for.

Gestalt-psychologists have also been accused of failing to make clear whether the *Gestalt* in any given case is shape, form, extent, direction, curvature or contour. This criticism involves the error of limiting the concept of *Gestalt* to the visual field. In fact, to serve as a comprehensive concept, this term should never be so regarded.

The last mentioned misconception is a natural one, for *Gestalt* is primarily a visual term. It is this visual connotation of the term that is apt to lead to much confusion of thought. To illustrate again, we may mention a criticism regarding the primacy of *Gestalt*, which reads as follows: "If the configuration means form proper, then we can say very definitely that animals perceive sizes and qualities before they do forms. If the primitive configurations are taken to mean figure-ground structures, according to Koffka, then the problem of genesis remains to be settled. If a configurational response means that an animal can react to a part of a situation in the light of the whole according to Köhler, then we have a different set of questions to answer." [3] Here, evidently, the critic is visualistically inclined in spite of the fact that he is quite well-read on the subject. This instance alone is sufficient to show the misleading influence of such terms as *Gestalt*, structure and configuration. F. Krueger [4] has also pointed out the source of this danger (although from another standpoint), namely that "the concepts '*Gestalt*' and '*Gestaltqualität*' have long been understood in the one-sided sense of a visual relationship."

Now, in the hope of eliminating this source of danger, I venture to suggest the term 'organic unit' as a substitute. The appropriateness of this new term will become apparent from the following considerations:

(1) The word 'organic' is defined by Webster as "possessed

[3] Helson, H., 'The Psychology of Gestalt,' *Amer. J. Psychol.*, 1926, 37, 210.
[4] Krueger, F., *Neue Psychol. Stud.*, 1926, 2, 3.

of a structure, or forming a totality, in which the relations of the parts involve relations to the whole, which itself is free or comparatively free from external relationship." The same author also defines the word as "forming a complex, self-determining unity." These notions are exactly what characterize a *Gestalt*. As to the idea of wholeness, it is adequately expressed by the word 'unit.' In this connection, I should mention the fact that the word 'unit' has already been used again and again by Köhler in his lecture entitled 'An Aspect of *Gestalt*-Psychology.'

(2) Being without necessarily visual implications, the proposed term is applicable to all forms of experience as a matter of course.

(3) The use of the new term would facilitate the attempt to distinguish between the qualities that are common to all forms of experience and those that are peculiar to particular forms of experience. In other words, the confusion of visual with other forms of experience would be thereby avoided.

(4) It should be understood that I should have proposed the term 'organic whole' as a substitute for 'Gestalt' were it not for the fact that there is no verb for the word 'whole.' This defect undoubtedly makes it much less fit to serve as such a substitute than the term 'organic unit.'

In view of the present situation of the new psychology, I also venture to offer the following suggestions for consideration:

1. Those who have gone through all the literature of *Gestalt*-psychology can never fail to realize the fact that, with the exception of Krueger's school, the study of emotion has been sadly neglected by the leading group of *Gestalt*-psychologists. Köhler has made only a few sporadic observations of emotional behavior; Koffka has touched only very briefly upon this topic in his writings; and none of this school have ever made a special study of emotion. It does seem imperative that more attention be given to this phase of behavior.

2. A clear distinction ought to be made between general and specific laws of *Gestalt*. The former refer to those that

govern all forms of experience, while the latter refer to those that govern particular forms of experience. In other words, there are laws which hold only within a particular sense-modality, and there are also laws which govern all *Gestalten* regardless of the differences of sense-modalities. In the *Gestalt* literature, there is no attempt to distinguish between the laws of these two classes. Through this lack of differentiation, they are naturally liable to confusion. In addition to this, there is a tendency often present on the part of *Gestalt*-psychologists to regard the findings in one field or in one case as true of all the other fields or cases. Such a tendency is apt to be misleading. To remedy this situation, our attention should be directed to the two-fold task, viz., the determination of the laws of *Gestalt* rather than the proving of its reality, and the differentiation between its general and specific laws. This would not only remove many an ambiguity in the *Gestalt-theorie*, but also open some new vistas in the field of experimentation.

3. The discovery of insight learning in animals is undoubtedly one of Köhler's valuable contributions to psychology; but what is more important is that, instead of merely determining whether or not a given case of learning involves insight, we should endeavor to discover both the general and the specific conditions that facilitate insight and those that inhibit it. It is in this direction, I believe, that research efforts will be richly compensated.

4. Lastly, I should like to call attention to the fact that many of the qualitative results of Gestalt psychology need quantitative consideration. An instance in point is Köhler's discovery that chimpanzees responded to the relation of things rather than to things themselves. Important as it is, this discovery still remains as a merely 'half-baked' truth. In order to make this contribution of real value to psychology, the statistical and developmental aspects of the problem have to be attacked. In fact, this need of careful statistical consideration is keenly felt in many *Gestalt* experiments.

V. Summary

In the introduction, I pointed out the necessity of representing the views of *Gestalt*-psychology owing to the nature of the English literature on the subject. It is with this fact in mind that this paper is written.

The first attempt of this paper is to clarify the concept of *Gestalt*. This is done by presenting its various aspects instead of giving it a mere definition. Such concepts as wholeness, transposability, phenomenal pattern and figure and ground, such principles as closure, intimacy, organization and formation and the Law of Prägnanz have been separately discussed and simply illustrated.

This is followed by a reformulation of the various psychological processes in the light of *Gestalt-theorie*. The processes here treated are perception, attention, recognition, memory, motor acts, thought processes, instinct, intelligence, habit, and learning. In the discussion of these, only the main features are given so that they may stand out clearly before our mind's eye.

In the next section of the paper, important accusations against the new psychology have been critically examined, and a few suggestions have been offered for consideration.

It is hoped that, in spite of the numerous inevitable weaknesses of this presentation, the principal phases of the *Gestalt*-psychology may be hereby clarified, and that thought on this subject will be stimulated. If this much is accomplished, then the writer's efforts will not be in vain.

BIBLIOGRAPHY

Allport, G. W., 'The Standpoint of Gestalt Psychology,' *Psyche*, 1924, 4, 354–361.

Benussi, V., 'Versuche zur Analyse taktil erweckter Scheinbewegungen etc.,' *Arch. f. d. ges. Psychol.*, 1917, 36, 59–136.

Calkins, M. W., 'Critical Comments on the Gestalt-Theorie,' Psychol. Rev., 1926, 33, 135–158

Helson, H., 'The Psychology of Gestalt,' *Amer. J. Psychol.*, 1925, 36, 342–370, 498–519; and 1926, 37, 25–62, 189–216.

Herrmann, J., 'Gesamterlebnisse bei Gerüchen,' *Neue Psychol. Stud.*, 1926, 1, 473–506.

Kantor, J. R., 'The Significance of the Gestalt Conception in Psychology,' *J. Phil.*, 1925, 22, 234–240.

Köhler, W., 'Die physischen Gestalten in Ruhe und im stationären Zustand,' 1924.

Idem, 'The Mentality of Apes,' 1926.

KOFFKA, K., 'Beiträge zu Psychologie der Gestalt,' *Zsch. f. Psychol.*, 1913, **67**, 353–358.
IDEM, 'Perception: An Introduction to Gestalt-theorie,' PSYCHOL. BULL., 1922, **19**, 531–581.
IDEM, 'Psychical and Physical Structures,' *Psyche*, 1924, **5**, 80–85.
IDEM, 'The Growth of the Mind,' 1924.
IDEM, 'Zur Theorie der Erlebniswahrnehmung,' *Ann. d. Phil.*, 1922, **3**, 375–399.
KRUEGER, F., *Neue Psychol. Stud.*, 1926, **2**, 3.
MESSER, A., 'Hüsserl's Phänomenologie in ihrem Verhältnis zur Psychologie,' *Arch. f. d. ges. Psychol.*, 1911, **22**, 117–129.
PILLSBURY, W. B., 'Gestalt vs. Concept as a Principle of Explanation in Psychology,' *J. Abn. & Soc. Psychol.*, 1926, **21**, 14–18.
'Psychologies of 1925,' Clark University Press, 1926.
Psychol. Forsch., Aug. 1921–June 1927.
SPEARMAN, C., 'The New Psychology of Shape,' *Brit. J. Psychol.* (*Gen. Sect.*), 1925, **15**, 211–225.
WASHBURN, M. F., 'Gestalt Psychology and Motor Psychology,' *Amer. J. Psychol.*, 1926, **37**, 516–520.
WERTHEIMER, H., 'Experimentelle Studien über das Sehen von Bewegung,' *Zsch. f. Psychol.*, 1912, **61**, 161–278.
IDEM, 'Über Schlussprozesse im produktiven Denken,' 1920.

[MS. received January 16, 1928]

THE GESTALT ENIGMA

BY HORACE G. WYATT

University of California

Now that its exponents in this country are familiarizing us with the Gestalt Psychology, the time is perhaps ripe for suggesting questions to which a preliminary study of the doctrine gives rise. To seek further enlightenment seems at the moment particularly desirable, since there appear to be a good many, like the writer, to whom the theory is in some respects baffling, and who have yet to be convinced that their perplexities are due to their own want of insight or comprehension rather than to some inherent deficiencies of the doctrine itself. As one who does not profess to have surveyed the Gestalt position at all completely, but who is anxious like many others no doubt to correct his provisional picture, I propose in this article to risk conviction of ignorance or naïveté and to attempt to formulate as clearly as possible some of the queries which the doctrine suggests.

As against the mechanistic and associationist accounts of human behavior and development the Gestaltist observations and arguments appear convincing, though we must not expect the very articulate protagonists of the popular 'scientific' principles, especially those who are committed to particular 'isms,' to yield their positions without a continued struggle. The less so, as after all the Gestalt criticisms of the mechanists are not new, though they come with the prestige of a new and vigorous movement behind them.

But the attack has been strengthened by more than a new name and fresh leading. The Gestalt principle and the associationist principle are diametrically opposed; and after encountering the remarkable ubiquity of the Gestalt principle, detectable in every angle and corner of mental life, even in strange and unexpected places, the associationists can no longer retain a foothold. By pushing its way in everywhere

Gestalt has been squeezing mechanical association out of existence.

To destroy however is one thing; to construct is another. The mechanist association doctrine has at any rate one merit: it does attempt to explain the *development* of the individual. We are all familiar with the ingenious expedients of trial-and-error and the conditioned reflex. Set in motion by 'drives,' 'desires,' or 'urges' (there is variety of terminology), the form taken by certain 'prepotent reflexes,' these twin devices are claimed to account for all human learning. And they are simple and intelligible; and consequently popular with young university students. The only objection to them is that they do not do the work, as the Gestalt psychologists amongst others have shown us. But have the Gestaltists an alternative to offer? They have demolished man as a marionette, together with the machinery and the operator, and they have set up in its place man as a moving picture show. But the trouble is that they appear to have forgotten to provide an operator. At least that is the first of our perplexities.

There are in fact two questions to which a clear answer is invited:

(1) Who or what makes the pictures? and
(2) Who moves them? and How?

I am quite aware that the analogy of a moving picture show is a little misleading. Gestalt is not confined to the visual field. It is being repeatedly illustrated in every sense modality. Moreover it is exemplified in instinct, perception, memory, intelligence, motor skills, etc., indeed in every type of mental occurrence. But that is not the point of my analogy. Substitute for pictures melodies or patterns or configurations or shapes or structures or what you will. The difficulty about them is still to find a definite and consistent answer to the questions: What are the laws of their origin? and What are the laws of their succession and evolution in the individual? The importance of being sure of a clear and workable answer to these questions need not be demon-

strated. Unless we know the principles upon which in any individual case Gestalt A passes on to Gestalt B rather than to Gestalt C, D, or E, or any of an indefinite number of possible Gestalten, we can neither as theoretical psychologists explain, nor as practical psychologists (educationists for instance) direct, the development of individual intellect, personality, or character. The questions are all-important. Without an answer to them, the Gestalt psychology is merely a psychology of structure and adds nothing to our understanding of genesis. It is not a *dynamic* psychology at all. It becomes a descriptive and not an explanatory psychology.

Let us then consider what may be the answers to our leading questions. We can begin by confining ourselves to the second of them 'What governs the succession of the pictures?' One way of treating this inquiry is the nursery device of diverting the attention to something momentarily more attractive. Now no one would accuse the Gestaltist of deliberately practicing this simple fraud; but for all that we may pay the Gestalt principle the compliment of saying that its very attractiveness has this diverting effect upon the unwary student. He finds himself overlooking the really important question because he is absorbed in studying the attractive features of the Gestalt itself. He is lost as it were in admiration of it. For the concept of configuration or Gestalt elaborates prettily enough into the aspects of figure and ground, and of wholeness, and exhibits the principles of Prägnanz, of closure, and of inner unity. Each of these aspects deserves and claims his attention. They appear to characterize mental phenomena on every side; and it is an interesting and illuminating exercise to watch for their appearance in every nook and cranny no less than to see them reveal themselves on the common surface of daily life. Nor need the student experience much difficulty with these features of the Gestalt. On the contrary he extends them a welcome; since he finds or thinks that he finds old friends under the new guise. For traditional psychology has already taught him that it is a function of mind to discriminate and select from its data (the concept of figure and ground), to

unify experience (the concept of wholeness), to be ever reaching a higher discrimination level (Prägnanz or definition), to anticipate or preperceive (closure) and to interrelate presented data (Innigkeit). He is thus predisposed to regard the doctrine with favour, to be interested in its far-reaching applications, and to follow it into regions hitherto unexplored. Impressed by its apparent universality he is apt to credit it with more than it is able to accomplish, forgetting that the laws of the Gestalt as statements of these attractive facts of form or feature may still tell us nothing of the conditions of their birth. What principles determine the emergence or occurrence of one particular Gestalt rather than another? This question is not answered by dilating on the different attributes of the Gestalt itself.

There is, however, another way of dealing with this question. An answer to the question 'why one Gestalt and not another' is forthcoming in the case of the process known as learning, meaning by learning the more or less persistent attempt on the part of man or animal to solve a problem or attain a new achievement. Not all human development is covered by this kind of learning; but a very great deal of both intellectual and motor (as well indeed as moral) development is of this nature or depends on learning in the sense given. In arriving at his configuration in this case the animal or human is said to display 'insight.' To this the reader will probably assent: learning does progress through insight. But if it is to throw additional light upon the 'choice' of configurations, to explain how it comes about that one configuration emerges rather than another, it is not sufficient merely to refer us to 'insight.' To give the configuring faculty the name of insight adds nothing new. You do not explain the share of the mind in the Gestalt process by giving it a name; all you do is to make it easier to talk about. The inquirer must not be put off by a mere tautology.

This however brings us to a third way of answering our question. For it is clear that there is no such thing as insight 'in vacuo.' Insight after all is a relative term: there must be something there is insight into. In other words

there must be something other than the mind itself which contributes to the insightful experience. It is obviously of first importance to know precisely what this is, for here we are after that something, external to the subject configuring, which provides a condition of his particular configuration. What then is it? A clear and unequivocal answer to this question will give us the essential cue. Now it is unfortunately just at this point that the doctrine ceases to be explicit. In discussing the phenomenon of learning (in the sense given) we are referred for an explanation of the Gestalt to the 'situation' in some instances and to 'the felt need' or the 'goal' in others. The ape for example is said to reconstruct the situation. There is a transformation of the field which precedes the final solution. And so on.

But what we require in this type of explanation is to know exactly where we are. What for example is a 'situation'? So far as a 'situation' helps to determine a configuration, it must surely mean one of two things. Either it is the outside environment, the objects in the perceptual field, and they constitute the influences or the stimuli which provoke the Gestalt to arise, or it is not. If it is this which is meant, then we are back again at determinism, since the configurations are determined by the environment. This however is the very doctrine which the Gestalt psychology condemns. If on the other hand the 'situation' includes the mind's interpretation of the environment, then we require to know why it interprets one way rather than another. The answer to this question is given where it is given at all, in terms of 'need' or 'goal.' There is some end or purpose, the impulse or desire to fulfill which, determines the particular way in which the perceptual field is reconstructed or patterned. Now in particular cases this is by the Gestalt advocate actually admitted. But to admit purpose is to dethrone the Gestalt. So soon as a 'situation,' in the sense of an *interpreted* situation, interpreted according to need or end, and so soon as need or goal are allowed to contribute to the configuration, the Gestalt ceases to exist in its own rights. It is no longer the Gestaltedness of the Gestalt that is the important thing

in the development of the individual but the extent to which it fits a 'situation' or fulfills a need or serves the attainment of a goal. We are thus back again at the traditional psychology of purpose; and the distinctiveness of the Gestalt psychology disappears.

Gestalt, we are forced to conclude, owes its 'Gestaltedness' to some non-Gestalt principle; it is not the Gestalt which is the significant and explanatory principle in developmental psychology, that gives to mental events, at least in the case of learning, their organic unity, but the existence and entertainment of ends and purposes. That this is in effect admitted by the Gestalt exponent himself is shown by the distinction that is drawn between a correct and an incorrect configuration, and between a 'good' and a 'bad' error. The ape's configuration is correct or his error is a 'good' one in cases in which it is the whole or part of a performance unified to effect his purpose or carry out his end. There is obviously no special appropriateness in a configuration as such. It may be a 'right' or a 'wrong' configuration. If it fits the 'situation' that serves the fulfillment of the end or purpose, it is a right configuration; if it does not, it is a wrong one. Configuration *qua* configuration is therefore no solution of the problem of development or of learning. It is not an ultimate but a secondary principle of behavior. It is in fact merely an interesting metaphor or analogy for the way the mind shapes its data or material so as most effectually to carry out its purposes.

Now this way of stating the matter throws a real light on the nature of intelligence. The insight of the Gestalt is no longer a mysterious power of forming configurations, a 'Gestalt' capacity. It is not the mere shaping of situations that constitutes 'insight' any longer. Its essence lies in shaping according to end or purpose.

But if this is the case, is this shaping, or is the shapeliness of the shapes, really so very important after all? To raise Gestalt to the dignity of a prime principle is to divert attention from the facts of most significance. In the actual process of learning the important things to get hold of are the facts of

the *process*, the process of *arriving* at the achieved Gestalt before we arrive there. In some cases no doubt we arrive at a solution in a leap, as on several occasions did Köhler's apes. But more often we do not do so. We just as often put our configuration together, as it were, piecemeal. Here a little and there a little. And sometimes not till long afterwards do we attain, if indeed at all, the final form. If I am learning so simple an operation as sharpening a pencil with a penknife, I find myself apprehending first one relation between my movement and the end and then another. I learn for example that pressing lightly gets a better point than pressing hard; then that a certain even stroke is the more effectual; that it is necessary to give separate attention to the leaden point and so on. In this way I gradually build up a configuration bit by bit. But in every step be it noted it is not the object before me, not the objective situation, but my conscious purpose, my end in view, that directs and 'chooses' the relations which I apprehend, and the consequent performance. Or to put it more exactly (so as to avoid the fallacy of hypostatization) I apprehend and select relevant relations in the light of my purpose or end.

To admit this point is to require a redefinition of the term Gestalt. For if the members of the configuration, the different parts of the completed performance, are not there, and of the kind they are, merely in virtue of their membership of a configuration, then Koffka's definition of a configuration appears a misleading one[1]—"a coëxistence of phenomena in which each member carries every other, and in which each member possesses its peculiarity only by virtue of its connection with all the others." This definition seems misleading because it excludes all other determinants of the character of the configuration and of its constituent members than the configuration itself. A more correct definition would seem to be, so far as applies to human learning at any rate, "a configuration is a coëxistence or sequence of phenomena in which each member is what it is and where it is in the whole

[1] Koffka, K., 'The Growth of the Mind,' N. Y., Harcourt, Brace and Co., 1925, p. 131.

pattern or performance because of the relation of the whole to some end entertained by the organism *other than the configuration itself*." In simple language there is a designer that determines the design.

We have now considered three ways of finding within the limits of the Gestalt purview an answer to our question 'What determines the sequence of the pictures'? What determines their succession and progress? And we have found these answers unsatisfactory. There is a fourth way of meeting the question. It consists in regarding it as superfluous, or at any rate in giving it up. This is in effect the course followed when it is maintained, as it is maintained, that the Gestalt is an ultimate unit of mental structure, a unit that does not admit of further analysis. If this is meant to apply to all Gestalten, the history of the individual's development resolves itself into a succession of mysteries, each a separate bolt from the blue. If it merely refers to a set of original or connate Gestalten, then we are back again at the problem of how the earlier Gestalten lead on to the later. The statement that Gestalten are ultimate gives no help here.

Another possible answer is given in terms of the increasing complexity of the Gestalten.[2] "The nature of mental development" we are told, "is not the bringing together of separate elements, but the arousal and perfection of more and more complicated configurations in which both the phenomena of consciousness and the functions of the organism go hand in hand." To this the rejoinder would seem to be that though it is no doubt true that as we develop our achieved Gestalten do exhibit increasing complexity, the direction which they take cannot be decided by complexity as such. For on any occasion there are many possible complex configurations open, so to speak, to the organism to achieve; so that we are left with the question which of them is to be selected, and why. Nor is the decision always in favor of the most complex, nor yet of the next most complex of all that are possible. In fact the complexity is not the determinant but the result of

[2] Koffka, K., *op. cit.*, p. 356.

the development. The fallacy consists in mistaking an incidental result for an efficient cause.

We may now return for a moment to the first of our two questions, the question namely 'Who or what makes the pictures?' Perhaps at any rate we may find an unambiguous answer to this elementary question. 'How much,' we may ask for example, 'is contributed by the organism and how much by the environment?' When however we turn to the literature for a reply we find the Gestaltists themselves in disagreement. There appear to be two schools of opinion. One of them holds that a Gestalt or at least the original Gestalten are already given in sensation: they arrive at the mind so to speak ready-made. The other school thinks otherwise. The interpretation of the sense data on the part of the mind is the first Gestalt. Some one has christened the rival schools the one-steppers and the two-steppers, for in the one case the Gestalt is immediate and in the other it is the result of a two-step process, sensation plus a creative act of the mind. To the student of genetic psychology seeking some guidance for practical application to know the answer to the question is not unimportant, for the reason that if the one-steppers are right the burden of responsibility for the individual's development will fall on the environment; so much will depend upon the sense data provided. He must shape the environment accordingly. Whereas if the mind creates the Gestalt, then the responsibility lies with the subject; and we must direct our efforts to the shaping of his mind.

However that may be, enough has been said to show the kind of perplexities that beset the naïve inquirer seeking from the Gestalt psychology some solution of his problems of the development of individual behavior. He may be pardoned if he finds himself regarding the evolution of the human individual as if he were viewing a kaleidoscope or a set of dissolving views; with the Gestalt guide at his side pointing out their different attractions but tantalizing him with mystifying answers when he asks how the thing is worked. He admits that the Gestalten shown to him are many of them

interesting, even remarkable. But he appears to be an onlooker at a new and fascinating game. The game of psychology seems to consist in discovering as many Gestalten as possible, and in the most unlikely and unexpected places. The discoverer of the greatest number in the strangest places wins the game. The rather special selected and laboratory character of the examples chosen to illustrate the Gestalt phenomenon also proves disconcerting, because it sets him wondering why if all mental life functions on Gestalt principles, more trouble is not taken to derive the principles and to display their features from the common stock of ordinary experience. That however is probably hyper-critical. It is his business as a non-laboratory psychologist to detect the Gestalten in ordinary life.

But there remains a further curious omission from the Gestaltic scheme of human growth and development. It appears to give inadequate consideration to the non-cognitional factors; especially to attention, emotion, and volition. Attention is admitted, so to put it, with the volitional element left out. And thus reduced its residual functions are transferred to the Gestalt. Volition however is not so easily disposed of. What account, it may well be asked, do the Gestaltists propose to give of it? For indeed in the development of the human individual there is perhaps no factor not excluding intelligence itself which takes a more important place than one's own volition. The formation of character in the sense of persistence in principle is conditional on acts of volition: the repetition of volitional acts is the foundation of character. And we all are familiar in our own experiences with the peculiar marks of volition. It consists, as we all know, at its highest, and distinctively human, level in the self inhibiting or controlling impulse. When we are aware of this unique phenomenon we are conscious of volition. It is not, of course, as is sometimes represented, merely the concluding stages of a conflict between two or more opposing impulses, for the conflict in the case of volition is not one between impulses at all. It is of a completely different order, namely, You or I imposing a check upon impulse, whether

that impulse be connate or acquired. Volition is the overriding of impulse. From the dominance of volition emerges character. To such as have not already committed themselves to some other psychological 'ism' the statement is a truism. But what place can Gestalt find for volition? None: for to admit this simple human phenomenon is to depose Gestalt. It is to let in a counter, and more than a counter, an overriding, principle of human development. Like his opponent, the mechanist, the Gestaltist has before him the problem of either explaining volition, that is finding a place for it in his psychological system, or if it has no place there, then of finding a place for the illusion. This problem, so far as I understand his position, he is content to ignore; an omission the seriousness of which it is impossible to overestimate, for on the proper exercise of volition more than on anything else the development of the human individual turns.

On the other hand the Gestalt psychology is not singular in this respect; so that we have no right to single it out on this count for condemnation. The tendency to ignore or belittle volition is the result of a misconception of the term science in relation to psychology. I do not propose however to discuss this question here. The point of immediate interest is rather the question of the mechanism by which so much modern psychology contrives to omit volition without the student noticing it. Indeed, one is inclined to suspect, without the champion of a particular psychological 'ism' noticing his omission himself. The thing is done by a skillful employment of the fallacy of hypostatization, or, as it is otherwise called, of misplaced concreteness. It is also known as the fallacy of reification, but even deification is not too strong a term for it. When dealing with discarded psychologies we are easily able to detect it. Herbart for example succeeded in giving a plausible air to his theories only by personifying or deifying 'ideas.' That is to say he seized upon a mere abstraction and endowed it with independent life, including among its personal properties will or initiative of its own. Nearer our own time we have the same fallacy

in Thorndike's 'Neurons' capable for instance of satisfaction and annoyance, and accounting on their own responsibility for man's behavior. By a similar expedient Alexander Shand succeeds in making 'Sentiments' responsible for the development of human character.

But can we exculpate Gestalt from the same fallacy of hypostatization? Consider the use made of the concept to explain intelligence. One of the criteria of intelligence is the manipulation of things with reference to the Gestalt in which they stand to one another. If for example you have four tubes of varying diameters, such that each of the smaller three fits exactly into the one next wider than it, and if they are laid in a row of increasing diameter before the ape under study, the ape shows his insight by picking up the tube with the hand nearest to the tube into which it fits exactly. He is intelligent if he handles the objects in accordance with the Gestalt in which they stand to one another. The fallacy here is the fallacy of hypostatization. For of course there is no Gestalt in which the tubes stand to one another apart from the mind imposing it. Similarly with the case of Attention. Attention from the Gestalt standpoint "is neither a mysterious faculty which clarifies the content of experience, nor an independently variable aspect of the phenomenal contents of perception; instead, attention is synonymous with the perception of a 'good' figure, the structure of which is articulate and definite. . . . Attention is only another word for the phenomenal definition of the contour and inner articulateness of a perceived figure." In other words it is the Gestalt which controls attention and not attention which controls the Gestalt. The Gestalt is thus given charge of mental activity, and attention becomes merely the experiencing of a Gestalt.

There is however no need to multiply instances of the hypostatization fallacy; for they are all referable to the initial fallacy of treating the problem to be solved as its own solution. Gestalt, one interesting aspect of organized experience, is abstracted from its concrete embodiments and given independent authority. Once you have done this all is

plain going: the Gestalt becomes an explanatory, a causative principle. We have endowed a mere abstraction with activity and initiative, and are saved the trouble of pushing the inquiry beyond the Gestalt itself.

By this convenient misconfiguration the actual facts of the situation are concealed, which are that in the concept of Gestalt we have no doubt a more accurate picture than the Associationists have painted of the actual structure of mental events, but that it is no explanation of the occurrence of a picture merely to indicate its accuracy. It does not tell us how the pictures come about, nor how they occasion other mental experiences and one another. Thus though the Gestalt as a descriptive metaphor may or may not be acceptable, it can be raised to the dignity of an explanatory principle only by illegitimately endowing it with the attributes of a person.

The student who is interested in the problem of the direction of human behavior towards a desirable goal emerges from his study of Gestalt psychology like some traveller upon a mountain top, from which he can survey (with field glasses and telescope) the many and varied scenes below. Somewhere away in the distance he can dimly descry the New Jerusalem, but when he consults the map that his guide has provided him he finds the directions for the journey have been left out. And the country for all its beauty spots has no signposts and no roads. He resolves to return the map with the request that the roads may be inserted.

[MS. received January 9, 1928]

AN ECLECTIC THEORY OF VISION[1]

BY GILBERT J. RICH

Chicago, Illinois

A theory of vision, like a theory in any other sense department, aims to explain the facts of sensation by a hypothetical mechanism in the sense-organ or sensory center. The value of such a theory must always be measured in terms of its adequacy as a means of explanation of the observed facts. No theory is to be considered as true or false so long as it remains a theory. If it can be proven to be true it graduates from the plane of hypothesis to that of fact; while if proven false it becomes an interesting item in the history of scientific thought. But it may properly be asked whether or not a theory is adequate to explain the known psychological facts of the sense-department in question.

If the field of visual theory is viewed from this standpoint, it is obvious that no one of the three leading theories, the Young-Helmholtz, the Hering-Müller or the Ladd-Franklin, offers an adequate explanation of even the principal facts of vision. Where one adequately explains, another totally fails. Of the Young-Helmholtz theory little need be said. Despite its wide acceptance among physicists and physiologists it really explains little more than the facts of color-mixture, and it can hardly be said to be considered seriously among present-day psychologists.

The theory of Hering, with the addition suggested by Müller, has the great positive merit of offering a logical and workable account of the colorless sensations. This result is accomplished by considering black and white as a disappearing color pair in every way analogous to blue and yellow or red and green, and by placing gray instead of black at the center of the visual system. The explanation of color, however, is

[1] Publication from the Institute for Juvenile Research, Chicago, Herman M. Adler, M.D., Director. Series C. No. 133.

not satisfactory. The justification of a tetrachromatic theory lies in the observed fact that there are four unitary experiences of color, the so-called 'principal colors.' But two of them, red and green, are neither complementaries, after-images of each other, the contrast colors of each other, nor constant in indirect vision. They cannot therefore be sensations from unitary psychophysical processes of the type described by Hering, and it becomes necessary to postulate two 'basic colors,' bluish-red and bluish-green. Since these latter are no longer unitary sensations, the whole purpose of a tetrachromatic theory fails and the hypothesis of Hering must be considered inadequate to explain all the facts.

The developmental theory of Ladd-Franklin, on the other hand, overcomes the difficulty just mentioned by describing a mechanism such that red and green, obviously a disappearing color pair, give yellow when in combination instead of a colorless sensation. The 'principal colors' are then also 'basic colors.' This theory offers, perhaps, the most adequate explanation of the sensations of color that we have. But it fails woefully when applied to the black-gray-white series. The concept [2] that black is a constant sensation, cortical or retinal, which occurs when there is no light stimulating the retina is pitifully inadequate to the facts. We have shown elsewhere [3] that such phenomena of colorless sensations as contrast, adaptation, the negative after-image and the nature of the black-gray-white series can only be accounted for if black and white are considered as a disappearing color pair and explained by mechanisms similar to those for blue and yellow or red and green. Venable [4] has pointed out other

[2] Ladd-Franklin, C., 'Tetrachromatic Vision and the Development Theory of Color,' *Science*, 1922, 55, p. 560.

Neifield, M. R., 'The Ladd-Franklin Theory of the Black Sensation,' PSYCHOL. REV., 1924, 31, pp. 498–502.

Michaels, G. M., 'Black: A Non-Light Sensation,' PSYCHOL. REV., 1925, 32, pp. 248–250.

Ladd-Franklin, C., 'The Theory of Blackness,' *Amer. J. Physiol. Opt.*, 1925, 6, pp. 453–454.

[3] Rich, G. J., 'Black and Gray in Visual Theory,' *Amer. J. Psychol.*, 1926, 37, pp. 123–128.

[4] Venable, W. M., 'The Ladd-Franklin Theory of Color Vision,' *Amer. J. Physiol. Opt.*, 1925, 6, pp. 521–526.

difficulties in Dr. Ladd-Franklin's theory of black. He notes that combinations of black with a color exist; that the study of threshold sensitivity indicates that black and white are a disappearing color pair; that black is not aroused by zero stimulation but only by contrast with light; and that the electrical potential of the optic nerve, which is above that of the surrounding tissue when the retina is stimulated by light, falls after the light is shut off and then returns to normal after an interval.

In holding that black cannot be put into the same category as other visual sensations because it cannot be made to change intensity without changing quality, Dr. Ladd-Franklin falls into the error of confusing intensity of stimulus and intensity of sensation. In other sense departments these vary concomitantly. Not so in vision. Here gray must be considered as the minimum of sensory intensity, so that change in the direction of any color or of black or white is accompanied by an increase in intensity.[5] She also falls in error in considering gray as a black-white, while in reality it is a separate quality which is intermediate between black and white only in the same sense that white is intermediate between blue and yellow or yellow intermediate between red and green.[6]

Our brief survey of the most important visual theories has revealed that one of them explains colorless sensations in an adequate manner, while the other offers a satisfactory account of the psychophysical processes in chromatic vision. It would seem, then, that a theory combining the advantageous features of both might offer hope of including in its explanations most, if not all, of the major facts of vision. The hypothesis to be outlined below attempts to accomplish this result. Our construction is to be built upon the principles that gray is the center of the visual system and the minimum of visual intensity, and that black and white constitute a

[5] Titchener, E. B., 'Visual Intensity,' *Amer. J. Psychol.*, 1925, 34, pp. 310–311.

[6] Dimmick, F. L., 'A Note on the Series of Blacks, Grays, and Whites,' *Amer. J. Psychol.*, 1920, 31, pp. 301–302.

Rich, G. J., *op. cit.*

Dimmick, F. L., 'The Series of Blacks, Grays, and Whites,' PSYCHOL. REV., 1925, 32, pp. 334–336.

disappearing pair. With such changes as will embody these points the developmental schema of Ladd-Franklin is to be utilized as the most satisfactory explanation of chromatic phenomena.

We may start our genetic schema with the least differentiated response to stimulation by light waves which we know—gray. The type of response represented by gray is the simplest visual experience, since it is differentiated with respect neither to intensity nor to wave-length. Whether or not it is the primordial sensory stuff which is experienced by organisms with an eye simpler than that of man [7] is only a matter of speculation. But the fact that gray is the sensation resulting from stimulation by heterogeneous light of moderate intensity indicates that it is the most primitive of the sensory processes of the human eye. Gray is to be explained by the hypothesis of a photo-chemical substance or color molecule (the two terms are synonymous at this stage) which is sensitive to light in the same way that the substances contained in the cells of other receptors are sensitive to their particular type of adequate stimulation. This molecule mediates but one type of sensory response—grayness, which has but one quality and which lacks such phenomena as contrast and negative after-image. In the human eye it is found in its undifferentiated state in the rods, whence comes the gray vision of low illumination. For in night vision only gray is seen, and there is no black, after adaptation excludes the more highly differentiated reactions to be considered below. Since this molecule is to be found in the rods, it serves to explain the entire range of phenomena of twilight vision which were dealt with by von Kries in his theory of dual vision.

At the next stage in the development of the visual process there comes a marked change. The sense of vision as it exists in man is different from the other senses in two respects. On the one hand, it is ever present. There is no visual zero, but only the visual minimum of gray. More important, however, is the fact that every phase of the sensory process except gray

[7] Ladd-Franklin, C., 'The Physicist and the Facts of Color,' *Science*, 1927, 66, p. 591.

is connected with a negative phase which occurs when or where the particular stimulus ceases to be effective, as in the phenomena of adaptation, negative after-image, and contrast. This negative phase is conceived by Ladd-Franklin as due to the action of the complementary part of the dissociated color molecule. In order to account for the achromatic sensations of black and white, the phenomena of negative phase must be explained. But it has already been pointed out that black and white are to be considered as a disappearing pair, so that the principles of the genetic theory may be directly applied. We may suppose that the first step in the development of vision after the primary sensitivity to gray is a partial dissociation of the primitive molecule into two parts. One of these parts is decomposed by the action of light and mediates the experience of white. The other tends to decompose spontaneously when the white portion is split off and mediates the sensation of black. It accounts for the black after-image and contrast effect of white and for adaptation to light, since it forms with white a disappearing pair. The primary color molecule has now become differentiated only with respect to intensity and still responds to all wavelengths with the same response. As one might expect, the resulting experiences of black and white resemble the more primitive gray more than do the chromatic sensations which result from the later differentiations with respect to wavelength. It is on account of this low degree of differentiation of black and white from gray that the latter has been mistakenly called the intermediate between them. Color molecules which have reached only this stage of differentiation are to be found in the extreme periphery of the retina and the eyes of one type of color-blind individual. They, together with the two remaining stages of differentiation, are to be found in the cones of the retina.

At this stage of differentiation the constancy of visual experience is present and must be accounted for. In a prolonged absence of stimulation or in a balance of the effects of light and its immediate absence (as in a mixture of white and black) there is the constant but minimal visual

sensation of gray. Some constant visual process must be assumed, for one always sees something. It is a matter of little moment whether this constant gray is hypothetically located in the retina, the thalamus, or the cortex. For the retina contains both a series of sense organs and a brain center (being derived from an outpouching of the embryonic brain) and the optic nerve is in reality a brain tract. It would seem that the constant stimulation resulting from the so-called endogenous retinal processes is sufficient to account for all the phenomena for which a constant visual experience is necessary, and is a vastly simpler hypothesis than is a cortical gray. The endogenous retinal processes, being weak, stimulate only the primitive gray which now becomes the constant visual process.

It may be objected that, in postulating a splitting of the simple color molecule in two parts one of which reacts in light and the other in its absence, we have added a feature found nowhere else in a sensory process. This is no doubt true, but vision presents unique features which require explanatory hypothesis different from those of other sense departments. In no other sensory field is there a real negative after-image, a true simultaneous contrast effect, nor a mixture of two qualities to give an experience which is in no sense intermediate. Moreover, the developmental theory requires exactly this type of splitting in order to explain chromatic sensations. It postulates that the part of the color molecule left when blue is split off tends to decompose and give the sensation of yellow in the absence of any additional stimulation. We merely postulate a similar mechanism at a lower level and suppose that the portion of the molecule left when white is split tends to break down spontaneously and to give rise to the sensation of black. By so doing we provide a ready explanation for the facts of achromatic vision. Adaptation, after-image, contrast and mixture in the colorless series are explained by exactly the mechanisms postulated by Ladd-Franklin to explain these same phenomena in the chromatic series.

The remainder of the theory requires only a brief state-

ment, for it is merely the existing developmental theory added to the foregoing description. We have already seen that in the second stage of the genetic schema the color molecule is split in two parts which are differentially active in light and in its absence. In the third stage the differentiation is with respect to wave-length instead of intensity. The white sensitive substance, which in its undissociated state is activated by light of any wave-length, now splits into two parts which are sensitive to the two ends of the spectrum. One portion, which mediates the sensation blue, is active when stimulated by wave-lengths under 544 $\mu\mu$, while the other responds to the longer waves with the experience of yellow. As in the preceding level, when either part is split off the remainder of the color molecule tends to decompose and produce its specific sensory qualities.

Finally, the portion sensitive to wave-lengths over 544 $\mu\mu$, the yellow-producing part of the molecule, divides further into two still more specialized portions sensitive to wave-lengths above or below 578 $\mu\mu$. At the same time they become partially responsive to wave-lengths beyond the range to which their mother substance reacted, so that the green-producing part gives responses from 477 $\mu\mu$ to 578 $\mu\mu$, and the red-producing portion to waves above 578 $\mu\mu$ or below 477 $\mu\mu$ to the limits of vision. The relation between the split-off part of the molecule and the remainder is in this stage exactly what it was in the two preceding stages.

The last two phases of differentiation which have just been outlined and the manner in which they so adequately explain the facts of chromatic vision have been too clearly stated by Dr. Ladd-Franklin to require repetition here. Our chief interest has been so to expand the developmental theory, by the addition of concepts derived in the first instance from Hering and Müller, as to make it equally adequate to the achromatic series.

To sum up; we have postulated the following visual mechanisms in four developmental stages: (1) A primitive color molecule or photo-chemical substance without differential sensitivity either to intensity or to wave-length, which

mediates a gray. (2) Differentiation of this primary color molecule with respect to intensity into two portions, one of which is sensitive to light of any wave-length and mediates white, and the other of which is a split product that mediates black on spontaneous decomposition. (3) Differentiation of the white process with respect to wave-length into parts sensitive to short and long waves and mediating, respectively, blue and yellow. (4) Further differentiation of the yellow-sensitive part with respect to wave-length into portions mediating red and green. The whole follows a single schema and explains the phenomena of black and white by the same mechanisms that it does those of color.

[MS. received January 9, 1928]

ARE THERE ANY NATIVE EMOTIONS?

BY J. F. DASHIELL

University of North Carolina

The terms 'anger,' 'joy,' 'grief,' 'despair,' 'love,' 'fear,' 'ecstasy,' and the like have been current coin in literary exchange, both popular and technical, for centuries. Such names have been used over and over again as if they denoted distinct, discriminable functions in human psychology. Some authorities have conceived that the psychological functions denoted by some of these terms are derived from functions denoted by others; and the literature is full of attempts to compare and qualify these phenomena. Elaborate descriptions of a few of them have been offered by acute observers such as Lange, Darwin, and James, including analyses of the bodily changes involved therein. Theoretical considerations, however, have cast some doubt upon the accuracy of many such terms; and experimental work that has gone forward in recent years on different phases of the topic of emotion is about to force psychologists to face the fundamental problem and to reconsider the traditional conception of the type of behavior called emotion. It should be worth while for us to summarize some of the leading pieces of experimental research with a view to determining whither their results tend to point. For the purposes of the present paper, I shall restrict the term 'emotion' to the meaning 'visceral reaction pattern'; and the query I raise becomes this: On the basis of experimental research, are we warranted in assuming that there are distinct native visceral reaction patterns corresponding to the traditional names of the emotions?

I. Recent Experimental Work on Adults

Brunswick (3) has investigated whether there are distinguishing changes in gastro-intestinal tone involved in the emotional reactions to different types of stimulating con-

ditions. He used the inflated balloon technique with a balloon in the stomach and one in the colon, both registering changes of pressure by recording tambours pneumatically connected. To his subjects he presented such stimulating conditions as: pistol shot, water dashed into the face, malodorous material, a snake on the chest, and an electric shock. He obtained tracings showing changes in tonicity in the two regions; and these he compared with the reports of his subjects as to the kind of emotional excitement each had experienced. His results showed that there was some tendency for increase of tonicity to occur in reactions reported by the subjects as 'startle,' 'surprise,' 'disgust,' and a few others, and some tendency for decrease of tonicity to occur in what were called 'fear,' 'wonder,' 'tenseness,' and so on. Contradictory results, however, were obtained from different subjects, from the same subject, and even from the records of the stomach and of the colon respectively on the same subject at the same time. The conclusion was inescapable that, so far as concerns the tonic condition of the enteric canal, although changes in tonic condition do occur in emotional excitement, the character of the changes cannot be used as a basis for discriminating different kinds of emotions.

Work by Cannon (**5**) has long been familiar, and has been used by some as a basis for a grouping of the emotions. He investigated the rôle of the adrenal glands in the development of the excitement appearing in behavior denominated 'fear,' 'rage,' and 'pain'; and he elaborated a conception of the operation of the autonomic division of the nervous system. He held that when the visceral reactions are innervated via the thoracico-lumbar subdivision of the autonomic, behavior was of the emergency type and was to be observed in reactions known as 'fear,' 'rage,' and 'pain.' In contrast to this type of emotion are those visceral changes due to innervation over the cranial subdivision, producing emotional behavior that may be called of the 'hunger,' 'thirst,' 'comfortable,' 'easy-going' type; and, on the other hand, those due to innervation over the sacral subdivision producing

behavior that may be called 'sexual' and 'eliminative' types. Thus, a contrast is suggested between visceral operations underlying the 'emergency' emotions and the 'appetitive' emotions. Such a distinction is, however, not thoroughgoing; for conditions arise under which one of the subdivisions mentioned may coöperate with an antagonistic subdivision. Consider the fact that in the acme of sex excitement the thoracico-lumbar is thrown into action, and further the fact that intense 'fear' or 'anger' excitement involves activity of the bladder and colon produced by innervation over the sacral. It seems that Cannon's statement that 'the differential features of the emotions are not to be traced to the viscera' is strictly true.

Experimental work on emotional changes accompanying lying and deception has borne some fruit. With the sphygmomanometer Marston (**11**) found that when a subject was lying his blood pressure registered at least 8 mm. higher than when he was telling the truth. With the pneumograph Benussi (**1**) and Burtt (**4**) found that the ratio of time for inspiration to time for expiration was increased after lying more than after truth-telling. With modifications of the forms of instrumentation just mentioned, Larson (**10**) has had striking success in getting blood pressure and respiration changes with lying in cases of actual criminals. With the galvanometer, changes in bodily resistance to electric current are demonstrable as a phase of emotional excitement, especially when used with the word-association technique. Landis' (**8, 9**) critical experiments did not bear out the claims for detection of deception by blood pressure changes but did slightly bear out the claims for inspiration and expiration ratios. Now, so far as our present question is concerned, one thing is to be borne in mind throughout the consideration of all this type of experimental work: when the experimenters get evidence of visceral changes, they do not get evidence of *what kinds* of visceral changes they are.

Marston (**12**) has offered evidence for a differentiation of three kinds of emotion on the basis of blood pressure changes: 'anger' when there is an abrupt short-lasting rise, 'fear'

with a similar longer-lasting rise, and 'sex emotion' with a distinct drop. Blatz (**2**) used a falling chair as the stimulating situation and with the electrocardiograph and the pneumograph obtained graphs showing changes in heart action and in respiration immediately following the fall. In all subjects he reported a uniform heart change: a definite acceleration of the beat followed by a retardation, and an irregularity in rhythm; while in nine of eleven subjects he reported retarded rate and increased depth of breathing. This work of Marston and of Blatz has not been verified as yet by other investigators.

Several experimental studies have been devoted to the measurement of ability to name from photographs the sorts of emotions for which the original subjects had posed. Feleky, Langfeld, Allport, Ruckmick, and Gates have all reported striking ability in this respect. Landis (**6**, **7**), however, obtained results having a different bearing. He took motion pictures of subjects while a wide variety of emotional stimuli was being presented to them; and he found that there was little or no consistent correspondence between the motion pictures obtained and the subjects' report as to emotions they had experienced: no facial expression or group of expressions typified any named emotion in a consistent way. He suggests that the facial patterns are not expressions of distinct visceral patterns of emotion but are social expressions —forms of facial reaction that have become habitual through social conventions.

A word should be said concerning the great volume of experimental work that has been done in times past to determine what respiratory and circulatory changes are involved in emotional conditions as divided into 'pleasant' and 'unpleasant.' The instruments employed included pneumographs, sphygmographs, sphygmomanometers, and plethysmographs. From a great array of such studies, I have selected for mention a few leading ones and present them in the accompanying table. Inspection of it will reveal the fact that—whatever other use the terms may have—visceral reaction patterns cannot yet be differentiated as corresponding to the traditional terms 'pleasant' and 'unpleasant.'

CHANGES INVOLVED IN EMOTIONAL CONDITIONS DIVIDED INTO THE 'PLEASANT' (DENOTED BY P) AND THE 'UNPLEASANT' (U)

	Respiration		Circulation		
	Rate	Depth	Pulse Rate	Blood Pressure	Blood Volume
Mosso	.....	.....	.....	.....	P−
Lehmann	.....	.....	P− U+	P+ U−	P+ U−
Binet	P+ U+	P+ U+	P+ U+	.....	P− U−
Külpe	.....	.....	U−	.....	.....
Gent	.....	.....	P± U+	P+ U−	P+ U−
Zoneff and Meumann	P+ U−	P− U+	P− U+	.....	.....
Angell and Thompson	No consistent changes		No consistent changes		P− U−
Shepard	.....	.....	P− U−	P− U−	P− U−
Stevens	P− U−	.....	P− U−	.....	.....

It appears from this brief survey that experimentation has failed to date to establish describable differences of any stable and consistent form between visceral reaction patterns of adults. The question arises, are there any distinct patterns to be found in the visceral segments of the behavior of babies?

II. RECENT WORK ON INFANTS

The earlier critical observations of children and infants were refined by Watson and Morgan (**15**) when they systematically controlled stimuli applied to newborn subjects and noted the resulting behavior. Their findings are well known. To the stimulation, loud noise, being suddenly lowered, or being jarred, the babies responded with puckering and crying, catching the breath, and convulsive clutching; to which reaction these authors applied the name 'fear.' To the stimulation, hampering of movements, the reaction included stiffening, slashing movements, holding breath, and screaming; which reaction was named 'rage.' To the stimuli, patting, shaking, or excitation of sensitive spots, there followed the reactions gurgling, cessation of crying, and smiling; this whole response being called 'love.'

Now it is pertinent to note that Tolman (**14**) characterizes this distinction of the three emotions as having been based really not on the visceral segments of the responses but upon the gross behavior results—the character of the back-action-

upon-stimulus in each case. Thus, 'fear' is a protective response tending to avoid the stimulus; 'rage,' a protective response tending to destroy the stimulus; and 'love,' a response tending to continue and get more of the stimulus.

Recent work by Sherman (13) is in interesting contrast with that of Watson and Morgan. (1) To one group of observers he showed motion pictures of the stimulating circumstances (which included two used by Watson and Morgan) and the ensuing responses made by infants under 12 days of age, and the observers were asked to write the names of the emotions. (2) To another group were shown motion pictures of the responses only, the stimuli having been deleted, and they were asked to name the emotion. (3) For another group the stimuli and the responses were transposed in the film. (4) Other observers were seated before a screen, the infants were given their stimulations behind the screen, and the latter was immediately lifted. In general: (a) considerable confusion was apparent in the judgments made by graduate students in psychology, by medical students, and by nurses, pointing to the importance of the interest and attitudes of the observers; and (b) 'success' in correctly naming and differentiating between the different emotions was much greater when the stimuli were known than when not known, indicating that the judging was not purely a recognizing of *patterned reactions as such* but was guided in part by knowledge of the character of the stimulus. The subjoined table presents a few of the interesting irregularities in judgments. From such results it seems clear that the rôle of the scientific observer in determining the findings is very great, also that the naming of an emotion is determined much by the character of the stimulation observed—in a given sort of situation the child is *expected* to be 'angry,' to be 'frightened,' etc.; and the clear cut results of Watson and Morgan call for verification by other experimental researches. In other words, it would appear not to be established that the emotional reactions of infancy are discriminable as distinct visceral reaction patterns corresponding to traditional names.

JUDGMENTS BY GRADUATE STUDENTS IN PSYCHOLOGY

	When Methods of Stimulation were *Shown*				When Methods of Stimulation were *not* Shown			
	Delay in Feeding	Dropping	Restraint	Needle	Delay in Feeding	Dropping	Restraint	Needle
'Fear'.....	0	27	4	7	7	5	5*	9
'Anger'.....	14	4	24	13	13	14	13	8
'Pain'.....	2	2	1	13	3	3	4	2
'Hunger'.....	7	0	0	0	7	6	2	2

* When medical students and nurses were acting as judges, 'fear' was the name most frequently applied to the behavior consequent upon restraint when the restraint had not been seen by the judges.

III. Two Alternative Theoretical Assumptions

In the face of this situation we may elect either of two alternative explanations.

(1) One is that there *are* emotions corresponding to the different conventional names, but that the visceral core of each remains yet to be discovered.

(2) A very different assumption suggested is that such names as are conventionally used for different emotions refer to different types of viscerally facilitated or inhibited overt *behavior* patterns that have been classified and labeled more in terms of their *social* significance than in terms of their *visceral* components.

Further, so far as the visceral patterns are concerned, we may entertain the possibility that there are no *native* patterns of visceral reactions at all, but that the patterning that is to be later discovered in particular individuals is traceable to their experience, is acquired. Given at birth a repertoire of visceral reaction-elements, these may come to be integrated in the course of the individual's development.

But note that the environment directly controls much overt behavior and can ordinarily exert but *indirect* influence upon visceral reactions. In consequence, while we may expect to find some uniformity of gestural, facial, and other forms of expression in different individuals, and may further expect early habituation to the conventional social usage in

the applying of names thereto, still we need not expect to find much uniformity between individuals in the patterning of the visceral reactions acquired.

When two people manifest very similar behavior as respects attitudes set up toward their environment, the visceral segments of the behavior may show entirely different configurations. At the expense of over-simplifying, let me make the point clearer with a concrete example. When two different individuals attack, the overt behavior of the one may be supported by increase in the tonicity of one part of the alimentary canal, that of the other by increase in tonicity of another part; the behavior of the former may be facilitated by more rapid breathing or a changed distribution of blood, that of the latter by accelerated pulse and by increase of adrenal secretion.

On this conception, then, man is not born with definite patternings of his visceral responses worthy of being called distinct emotions. The most that he has in this direction are certain tissues, organs, and organ-systems (respiratory, circulatory, heat-regulating, etc.) capable of interacting and mutually influencing in complicated ways, and the rates, amounts, and orders in which these organs and systems severally participate in the overt behavior of a man toward his environment are not predetermined in any highly definite way.

Moreover, as these organs and systems come but little under the influence of those environmental controls that determine the integration of manual and thinking habits more or less similarly in all individuals, the particular combinations in which they come to function will show little or no uniformity from individual to individual. 'Emotional' and 'emotions' will still have validity and use in psychological description as generic terms, but 'emotions' as visceral pattern reactions may survive only as socially determined constructs.

References

1. Benussi, V., *Arch. f. d. ges. Psychol.*, 1914, 31, 244–273.
2. Blatz, W. E., J. Exper. Psychol., 1925, 8, 109–132.
3. Brunswick, D., *J. Comp. Psychol.*, 1924, 4, 19–79, 225–287.

4. BURTT, H. E., J. EXPER. PSYCHOL., 1921, 4, 1–23.
5. CANNON, W. B., *Amer. J. Psychol.*, 1914, 25, 256–282.
6. LANDIS, C., J. EXPER. PSYCHOL., 1924, 7, 325–341.
7. IDEM, *J. Comp. Psychol.*, 1924, 4, 447–509.
8. IDEM, *J. Comp. Psychol.*, 1925, 5, 221–253.
9. IDEM, *J. Comp. Psychol.*, 1926, 6, 1–19.
10. LARSON, J. A., J. EXPER. PSYCHOL., 1923, 6, 420–454.
11. MARSTON, W. M., J. EXPER. PSYCHOL., 1917, 2, 117–163.
12. IDEM, J. EXPER. PSYCHOL., 1923, 6, 387–419.
13. SHERMAN, M., *J. Comp. Psychol.*, 1927, 7, 265–284, 335–351.
14. TOLMAN, E. C., PSYCHOL. REV., 1923, 30, 217–227.
15. WATSON, J. B., AND MORGAN, J. J. B., *Amer. J. Psychol.*, 1917, 28, 163–174.

[MS. received January 23, 1928]

SOME FALLACIES UNDERLYING THE USE OF PSYCHOLOGICAL 'TESTS'[1]

BY H. M. JOHNSON

Mellon Institute of Industrial Research, University of Pittsburgh

It is said that in certain systems of Magic, if the name of one object can be securely bestowed upon another, the two objects thereupon become identical.

The rite which effects the attachment is intricate, and requires a capable wizard and a coöperative client to perform it. Still, if it be properly carried out, the results are miraculous. For example, if you desire to bring suffering and humiliation upon some enemy who is too powerful for ordinary tactics, you need only get his name transferred to a doll which you keep, and then maltreat the doll.

Miracles which depend upon this principle do not seem to work as well in Science as they are said to work in Magic. Scientific Method requires a distinction between *equivocance* and *equivalence*. For two objects to be equivocant they need only to bear the same name; for them to be equivalent (with respect to a particular function) they must be capable of interchange without altering any set of relations which are to be considered. The fallacy of treating two objects as equivalent because they bear the same name is called Equivocation. It is appropriate that this fallacy has an ethical aspect.

Although Psychology, in its efforts to become a Science, has formally repudiated Equivocation as a valid method, a great many psychologists are virtually trying to practice this item of the sorcerer's art. One of the most conspicuous examples lies in the usual mode of application of what have come to be known as psychological 'tests.'

As we are using the word 'test' a little more inclusively

[1] Presented in part before the American Psychological Association at Columbus, Ohio, December 29, 1927.

than is the present fashion, it may be well to pause an instant to define it.

In Science, a 'test' is an indirect means of measuring an empirical and variable quantity, which we may call Y, by measuring another variable, X, upon whose values the corresponding values of Y depend. For example, if one should attach a thermometer and a pressure-gauge to a closed vessel which contains a gas, and should make a series of simultaneous readings of the two instruments while the system was being heated and cooled, one would find that the temperature, x, could be used as a 'test' of the pressure, y, within a certain range, and subject to a margin of error, both of which can be computed. For, within this range, $y = f(x) \pm R$, in which R, the initial of the word 'residual,' represents the probable error of a single measurement. More specifically, $y = (k/v)x \pm R$. Until the first relationship has been restated specifically, X cannot be used as a 'test' of Y.

The special example illustrates the conditions which must be fulfilled before any X-variable can be used to test any Y-variable. First of all, both Y and X must be capable of being measured directly and independently. Secondly, the values of Y must depend on those of X in a manner which can be specified; otherwise there is no way of knowing *what* value of Y a given value of X indicates or 'means.' But when the results of a series of paired measurements are reduced to a descriptive equation, then Y can be tested by X, with a degree of precision which can be computed, and which serves to determine whether the test is or is not worth applying.

As a matter of fact, most physical measurements are made, not directly upon the variable under study, but rather upon some of its functions. For example, in as simple an operation as weighing by means of a steelyard, what is observed directly is the distance along the lever-arm at which a counterweight must be set in order to balance the mass to be 'tested.' The point to be remembered, however, is that this relationship between length and mass has been determined, at some

time or other, by some method of direct and simultaneous measurement on the two variables; until this was done the linear scale of the steelyard could not be calibrated in pounds.

But: suppose that Y is not amenable to direct observation and measurement. Does it then have existence other than the sort which one confers upon it by assumption? Can it be indirectly measured, or 'tested,' by X? These questions, though often raised, are absurd; for the operation which determines their answer cannot be performed. Scientific method prescribes their abandonment at this point.

But, how many of the variables about which psychologists talk, belong to the class of things which cannot be measured, cannot be observed, and should not be presumed to exist! 'Intelligence,' 'learning-capacity,' 'arithmetic ability,' 'interest,' 'intoxication,' 'fatigue,'—things which none can observe or measure, but for which they have been trying to find 'tests.' By what method have they hoped indirectly to measure the immeasurable? It is precisely by the method of Magic, or Sorcery, which we mentioned at the beginning. "Although we cannot observe or measure Y," they say, "let us give its name to some function of X. Let us now treat this function as though it were indeed the thing whose name we put upon it." The argument falls into this formal pattern: (I) By definition, $Y = F(X)$; (II) assume $Y = Y$; (III) therefore, $Y = F(X)$. The fallacy consists in the use of the assumption in (II) as though it describes any observable, measurable and objectively verifiable property of Y. The following illustration, expressed in verbal symbols, may suffice: Let us *define* 'general intelligence' (Y) as a function of the centile score (X) on the National Zeta test. Assume that 'general intelligence' (Y) as so defined is a measure of 'innate learning capacity' (Y), which is also called 'general intelligence.' It therefore follows that general intelligence, otherwise called innate learning capacity, varies according to the centile score on the National Zeta test. Q.E.D. Thus, the situation is made to appear as if the conditions of a genuine test have been met; whereas they have not.

Modern Psychology is often dated from Fechner's 'Ele-

mente der Psychophysik,' published in 1859. Fechner's system is founded on an equivocation, which persists as an essential part of every psychological system now extant, except behaviorism. Fechner's standard formula, $\gamma = k \log (\beta/b)$, defines a function of the ratio of the intensities β and b of two stimuli: nothing more. But upon this relationship Fechner bestowed the name "intensity of the sensation aroused by the stimulus whose intensity is β," as the savage bestows the name of his enemy upon his doll. Fechner had argued at length that a sensation cannot be measured even by its owner;[2] but after introducing his definition he ceased to question whether the new property he attributed to the sensation—*i.e.*, that of being measurable indirectly by the stimulus—belonged to it by assumption or in reality. It might have saved much stationery if he had reminded himself occasionally that no γ represented the actual intensity of a sensation, but merely the intensity which he *assumed* a sensation to have.

Following Fechner, who ought to be regarded as the first psychological tester, a host of other instances appear. For example: Can the depth of general sleep be measured by the depth of sleep with respect to a particular stimulus, granting that the latter can be derived from the threshold magnitude of the stimulus? Scientific Method asks, "Can the operation be performed which will answer the question"? Kohlschütter and his successors *assumed* an answer, which happened to be 'yes.' Is the variable they so *defined* the one they set out to *measure?* To satisfy the definition, one must say that under Kohlschütter's conditions, 92 per cent. of all the sleep taken occurred within the first two hours of the night. That is, unless one questions his data.

Can the degree of fatigue be 'tested' by impairment of performance in some task? If so, then a person may be more

[2] Much of what is currently said about Fechner is not based on first-hand acquaintance with his writings, but on accounts of those writings rendered by readers whose understanding of them is none too good. I wish that behaviorists in particular would read him carefully, for in his introductory chapters he states the argument for the behaviorist's position more aptly than any behaviorist has put it; although he presently bridges all difficulties by means of an *assumption.*

fatigued before he begins his day's work than when he is near its close.

Can recuperation from fatigue be tested by improvement in a task? If so, then one may be less recuperated after eight hours of sleep than before one went to bed.

Can the degree of alcoholic intoxication be tested by impairment in some performance? If so, under some conditions a highball may produce more intoxication before it is swallowed than afterwards.

Can a person's 'interest' in a lecture or conversation be measured by the time he refrains from 'blinking'?

Does the change in electrical resistance of the body which follows the propounding of a question indicate the kind and intensity of the emotion (introspectively observable) which the question excited?[3]

Is the diameter of the pupil, or rate of perspiration, or the electrical resistance of the body, a measure or test of the depth of sleep?

Has a child a certain fixed *capacity* for learning arithmetic, which is born with him, which can neither be increased nor diminished, and which can be measured by some standardized performance?

Can one tell how much pain an electric shock causes an earthworm, by measuring the time the worm consumes in separating itself from the electrode?

Can one test whether a dog, or an ape, has ideas, by noting whether it manifests behavior which we have decided to call ideational?

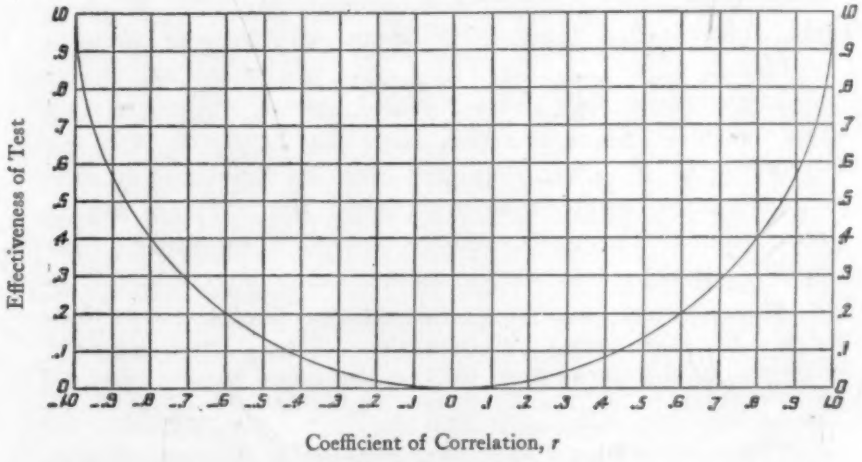
These are samples of current problems in psychology, all of which require one to find the relation between two variables, one of which is not, and cannot be, given. They are pseudo-problems, and they belong in a pseudoscience. One can lengthen the list almost indefinitely before having to name

[3] Incidentally this question may be a little further involved. The record of the galvanometer has two dimensions: extent and time. The emotion, by definition, has many dimensions. The absurdity involved in the problem is faintly suggested by the problem of the projection of solid objects on plane surfaces. Any number of such objects can be made to cast identical shadows. In brief, a variable cannot be adequately represented in fewer dimensions than it has.

the attempt to find a number which expresses 'How smart are your children?'

I now turn to another instance of the practical application of the fallacy, expressing the argument, as before, in mathematical symbols. An independently measurable variable, Z, is proposed for test by another variable, X. The mode and degree of interdependence can, of course, be ascertained by simultaneous measurements and mathematical treatment of the paired numbers which express them. This is the sort of situation which makes possible a genuine test. There are many such. But instead of finding the relationship directly, and expressing it directly, the following formal pattern is used: (I) $Z = f(Y)$, by admission; (II) assume $Y = \mathrm{Y}$, in which (III) by definition, $\mathrm{Y} = F(X)$; thence, (IV) $Z = \phi(X)$, Q.E.D.

As a particular illustration: (I) productivity in salesmanship (Z) is a function of 'social intelligence' (Y), as 'everybody will admit'; (II) this sort of 'social intelligence' (Y) is the same entity which the testers call *social intelligence* (Y);



(III) *social intelligence* (Y) may be defined as centile score (X) on Moss' 'test'; therefore this score (X) measures productivity in salesmanship. (These substitutions were made simply to illustrate the form of the argument.)

However, the effectiveness, E, with which any dependent

variable, such as Z, is indirectly measured, or 'tested,' by X, is given by the relation $E = 1 - s/\sigma$; and, in case the relationship is linear, by $E = 1 - \sqrt{1 - r^2}$, in which r signifies the coefficient of correlation.[4] From the accompanying curve (Fig. 1), it may be seen that unless the value of r is unusually large, E is questionably small.

In the argument just given the use of Y and Y as middle terms is confusing, as well as irrelevant. Its justification has been that it 'aids in selling the tests.' Inspect this curve,[5] recall that some of the tests sold in this manner are correlated with the trait by not more than 0.25 to 0.5, so that their effectiveness runs between 3 and 13 per cent., and the conviction may develop that the argument has sold some tests that would not sell on their merits.

Whatever the temporary effect may have been, the more lasting effects have been bad. Some of this work is now getting a valid appraisal; it is not astonishing that that appraisal is unfavorable. Nor is it remarkable or unfair that Psychology should be regarded as a pseudoscience, rather than as an embryonic science, as long as so much of her literature is devoted to sham solutions of pseudo-problems.

My concern is for the situation in which legitimate tests are possible. Those are the situations in which both the variables to be correlated are independently measurable. In such cases tests can be constructed whose values can be ascertained. The earlier efforts may not be immediately useful, but the results can be made clear.

[4] The symbol σ represents the standard deviation, or quadratic mean deviation, of the various empirical values of Z from their mean, $\bar{Z}$; *i.e.*, $\sigma = \sqrt{(1/n)\,\Sigma(Z - \bar{Z})^2}$. Similarly, s represents the quadratic mean of the deviations of the actual values of Z (such as $Z_1, Z_2, \cdots, Z_n$) from the corresponding values (such as $Z_1, Z_2, \cdots, Z_n$) computed from the regression equation $Z = \phi(X)$: *i.e.*, $s = \sqrt{(1/n)\,\Sigma(Z - Z)^2}$.

[5] The expression $E = 1 - \sqrt{1 - r^2}$ is the equation of a circle whose radius $= 1$ and the coördinates of whose center are 0, 1. Since the limits of E are 0 and 1, and the limits of r are 1 and -1, the points on half the circle satisfy all the real values of E and r. To plot the relationship quickly and accurately, it is necessary only to plot the two variables to the same scale, as in the figure, and describe a semicircle whose center is 0, 1 and whose radius is 1. If space must be conserved, as in a small notebook, the abscissæ may be used to represent absolute values of r; in which case a quadrant is adequate.

As an illustration I may give two replies to the question, "Why did you give that test, Professor"? It has drawn the answer, "Because it measures your general intelligence, from which I can estimate your academic achievement." A better answer is, "I shall use your score to estimate the success which you are going to show in 'getting by' the prof's. In its present form its effectiveness is *E*."

Addendum

This place may be suitable for statement of the only question which was raised at the time the foregoing paper was presented. That question was whether the point at issue was not largely a matter of diction rather than of logic. In general terms, cannot one give the name Y to a function of *X* without necessarily confusing Y with *Y*. For example, is it not legitimate to call a certain performance a 'test of intelligence'[6] if one uses the word 'intelligence' to cover only what the score measures? This is a stock argument and deserves special comment.

In the first and least important place, it may be replied that such reasoning is inconsistent with the definition of a 'test' as 'an indirect measure.' By substitution the proposition would read 'what this score measures indirectly is whatever it measures directly.'

Secondly, the fashion of naming 'test-performances' arbitrarily may introduce great confusion in case the same name should be applied to more than one performance. Suppose two performances are called 'tests of intelligence': shall we therefore assume that they 'measure' the same thing? If we did so, trouble would soon arise, for it may happen that the score in one performance is not a very effective index of the score in the other. A correlation of 0.8 between the two scores on two such performances has been considered high; but, as appears from Fig. 1, this coefficient indicates that either score is but 40 per cent. effective as a test of the other. Consequently the number of kinds of 'intelligence' tends to

[6] My critic used the word 'test of genius' rather than 'test of intelligence' but the form is identical for the two arguments.

approach the number of performances whose scores define them; to distinguish among these 'intelligences' we need as many names as there are defining performances. To add the designation 'intelligence' to each one merely increases the number of words to be printed and pronounced, *if* it is to mean no more than a function of score in some particular task.

The final objection may be phrased as a challenge. Can one produce a single paper in which the author arbitrarily defines a trait as whatever a certain performance-score measures, and then adheres to his definition? I have yet to find one, although I have not examined all the printed matter dealing with the subject, and shall not devote the remainder of my life to such a search. I am sure, however, that those school superintendents who criticize their teachers' 'methods,' particularly with respect to 'motivation,' on the ground that the pupils who get the highest marks in arithmetic are not invariably those whose score is highest in some 'test' of 'arithmetic ability' must tacitly assume that 'arithmetic ability' is something over and above a rating in some one specific performance. The official papers of the War Department do not indicate that the Alpha and Beta 'tests' were recommended for adoption on the ground that the score of each individual indicated merely what his score actually was. But I shall not burden this paper with illustrations.

In general, the clarity of a situation is always impaired if two non-equivalent things belonging to it are called by the same name; for by a process of training the equivalence of vocal responses aroused by the common symbol may become extended. The responses of a devout congregation to the consecrated Host are probably indistinguishable, in their essential features, from those which would be made to fragments of the Body of Christ if presented in its original form. This equivalence of response probably results from a process of training, in which an essential part is the oft-repeated declaration '*Hoc est enim corpus meum*'; and asseverations of the same tenor in the private prayers of the worshippers. Opponents of the doctrine of transubstantiation assert

(while its supporters deny) that the equivalence of the Host and of the original Body, is relative *only* to the responses of the faithful.

This dispute, considered as one of *fact*, has no scientific relevance; but its *form* may help suggest the rôle which Equivocation has played, and is playing in psychological method.

Useful References

1. Bridgman, P. W., 'The Logic of Modern Physics,' N. Y., Macmillan, 1927.
2. Keyser, C. J., 'Mathematical Philosophy,' N. Y., Dutton, 1922.
3. Lewis, G. N., 'The Anatomy of Science,' New Haven, Yale Univ. Press, 1926.
4. Poincaré, H., 'Science and Hypothesis,' N. Y., Scribners, 1914.
5. Yule, G. U., 'Introduction to the Theory of Statistics,' 7th ed., London, Griffin, 1924.

[MS. received January 18, 1928]

DISCUSSION

ON THE PRESUPPOSITIONS OF CHARACTER TESTING

The attempt to measure various character and personality traits, or a composite of these under the heading 'general character,' has in recent years become one of the most noteworthy developments in the field of mental testing. Opinions differ, however, as well they may at the present pioneer stage of this movement, concerning the value of what has been done. On the one hand, we have such optimistic conclusions as those of Haggerty, who, after stating that "Investigators in these matters have secured tests with coefficients of self-correlation ranging from 0.75 to 0.95, and correlations with other accepted criteria of the traits in question ranging from coefficients of 0.40 to 0.74," concludes that "This indicates that the measurement of some character traits has reached a state of reliability and validity comparable with the measurement of intelligence of a few years ago and equal, or superior to, many of the devices now used for measuring scholastic achievement." [1] On the other hand, we have the recent contribution of Witty and Lehman, who, beyond presenting what they conceive to be the difficulties of work in this field, appear to belittle the whole movement by the broad statement that, according to their views, "attempts at character measurement appear spurious and unnecessary." [2] They criticize attempts to measure not only general character but also honesty, trustworthiness and in fact any other trait which is ordinarily classified as a character trait. It seems imperative to examine in some detail the grounds for this critical attitude. Such an examination is undertaken here in the belief that it will reveal certain inadequacies in the criticisms which have been advanced, and in the hope that it may help to clarify the problems of character measurement and to remove possible misunderstandings which might interfere with further progress. The arguments and the viewpoints which will be set forth are definitely at variance with those of Witty and Lehman; but it is hoped that difference of opinion, however marked, will not

[1] Haggerty, M. E., 'Character Education and Scientific Method,' *J. Educ. Res.*, 1926, 13, 240–241.

[2] Witty, P. A., and Lehman, H. C., 'The So-Called "General Character Test",' Psychol. Rev., 1927, 34, 413.

suggest any lack of appreciation of the service of these writers in bringing together such a mass of criticisms as they present.

The fundamental reason why Witty and Lehman regard the attempt to measure character as 'spurious and unnecessary' appears to be that 'character is acquired in terms of *habits* of action.' The concluding paragraph of their paper is as follows: "If one accepts the point of view that character is acquired in terms of *habits* of action (and who has any evidence to the contrary?) the attempts at character measurement appear spurious and unnecessary." [3] Clearly, this statement implies that the attempt to measure any trait which is acquired in terms of habits of action is open to criticism. It appears, therefore, to constitute a condemnation of many measurements other than those of character, including some that have long been extensively used. Ability in the fundamentals of arithmetic, for example, is acquired in terms of *habits* of action, and yet most people regard it as desirable, or necessary for certain purposes, to attempt its measurement. Is the measurement of such ability also spurious and unnecessary?

In another place, it is stated that "General character appears to be largely the result of experience, not an innate trait. Until evidence is adduced to the contrary, it seems logical to accept this explanation. It therefore seems that attempts to measure general character are spurious when simple devices are applied which fail to take into account the varied experience of each child." [4] Do these statements imply, as further criticisms, first, that some of the tests are simple, and second, that they do not take into account differences between children in past experience? The first may readily be dismissed. It is no criticism of a test to urge merely that it is simple. The overstatement test, moreover, which is the object of special criticism, is certainly far from simple. The second criticism, that the tests do not "take into account the varied experience of each child" is not as precise as might be desired. If past experience is the chief factor in determining character, then individual differences in the scores of any character test must be due chiefly to differences in the past experiences of different individuals. The scores would then inevitably reflect, and in this sense, take into account, differences in past experience. There is no reason for disagreeing with Witty and Lehman in their view that the overstatement test is to a very considerable extent affected by such differences. These

[3] Witty, P. A., and Lehman, H. C., *op. cit.*, p. 413.
[4] Witty, P. A., and Lehman, H. C., *op. cit.*, p. 406.

writers point out how scores made in this test reflect differences in the classroom experiences of the children. "To take into account" past experience presumably does not, however, mean simply to allow differences in past experience to exert whatever influence they may upon the test scores. "To take into account" differences in past experience seems rather to suggest the idea of correcting for, or eliminating the effect of, differences in past experience. There is no insuperable difficulty in working out corrections for differences in certain aspects of past experience, and such work should be of value in the study of the causes of the trait measured. It should be remembered, however, that an allowance for differences in past experience, if perfectly carried out, would, to just the extent that differences in character are due to differences in past experience and are not innate, result in the elimination of differences in individual scores. If the scores are entirely a matter of past experience, elimination of the effect of past experience would result in all scores being identical. If character even in part is a product of experience, it is obvious that to eliminate the effect of experience upon our measurements of it, would be to falsify these measurements. A proposal for such elimination would be expected only from one who holds character to be some purely innate capacity, and not from one who emphasizes that character is derived from experience.

That there exists much misunderstanding concerning the relationship between the question of causative factors and the problem of measurement is evident from the erroneous view expressed by Witty and Lehman that Woodrow and Bemmels must consider character to be an innate capacity. The latter by no means consider character to be entirely innate; and as regards the extent to which it is innate, they expect to reserve judgment until presented with adequate data on the matter. No statement published by either of them appears to account for the assumption that they consider general character an innate capacity. Witty and Lehman write: "And they [Woodrow and Bemmels] must consider general character an innate capacity similar to the mental tester's conception of general intelligence since they assume character can be gauged at pre-school ages." [5] Since when, it may be asked, has the term, innate, come to mean present in children at the age of four? After all, mere opinions as to whether character is innate or acquired are of little importance. The point to be emphasized is that no assumption on this matter is a necessary prerequisite to the measurement of

[5] Witty, P. A., and Lehman, H. C., *op. cit.*, p. 406.

character. Once we have adequate measures of character, we shall be in a position to obtain decisive data concerning the relative importance of heredity and environment in its determination.

Another matter which needs clarification is the question whether an investigator, simply because he sets out to measure a trait of character, such as honesty, or a trait such as general character, is assuming that these traits are unitary. Witty and Lehman, who lay great stress on the view that character is a matter of specific habits, write as follows: "In the face of repeated protests a few experimenters still seem to be clinging tenaciously to the hypothesis that general character is a unitary trait, and that it may be measured by a simple device." [6] They add that the most recent attempt of this nature is that of Woodrow and Bemmels. It is impossible to understand how the latter could be accused of asserting or implying that character is a unitary trait. Does not their express statement that they understand by development in general character an average, namely, "the average development attained in all desirable specific traits of character" dispel the idea that they consider character a unitary trait? In order to make sure that it would do so, they added: "The number of these specific traits is very large. A list which includes only a small proportion of the total number of these traits, but which nevertheless embraces as many as two hundred of them, chosen as presumable constituents of that aspect of character called 'good citizenship' has been compiled by Upton and Chassell. Each of these hundreds of traits is presumably quite complex, like the traits of self-assurance and speed of decision, studied experimentally by Filter." [7] How, it may be asked, can one be accused of considering character a unitary trait when he defines it as an average of a *list* of traits so long that a list of two hundred traits constitutes but a small proportion of the total, and when, further, he expresses the conviction that each of the hundreds of traits making up this list is itself quite complex? Apparently one is to find the answer to this question in the following summary by Woodrow and Bemmels: "The overstatement test, so far as may be judged from our small experiment, constitutes a very good test of general character in pre-school children. We do not mean by this that taken alone it is an adequate test, but that it would not take a large battery of tests as good as this one (providing the intercorrelation between them was moderate) to yield a very satisfactory

[6] Witty, P. A., and Lehman, H. C., *op. cit.*, p. 405.

[7] Woodrow, H., and Bemmels, V., 'Overstatement as a Test of General Character in Pre-School Children,' *J. Educ. Psychol.*, 1927, 18, 239.

measure." It should be clear from the above quotation that the overstatement test by itself is not regarded as an adequate or valid test of general character. It is indicated that the overstatement test is a good test for use in a battery of tests for the measurement of general character. A criticism of these conclusions on the grounds that they place too high an estimate on the value of the overstatement test would be entirely legitimate; but the inference that they imply the view that character is 'unitary' appears to reveal some serious misunderstanding.

Does, then, the opinion that a satisfactory measure of character may be obtained from a suitable battery of tests necessarily imply a belief that character is unitary? It is hard to see how anyone familiar with mental measurements and the theory of correlation could answer this question in the affirmative. It is well established that desirable traits tend to be positively correlated. Traits may be more or less loosely connected, but they are not unconnected. Measurements of any one will show a correlation with many others. As a consequence, a serviceable measure of a vast number of specific traits may be secured by the use of a relatively small set of tests, providing these tests are so chosen that each of them correlates highly with many other tests and that the tests chosen do not correlate too highly with each other. It is not at all out of the question to obtain a fairly valid measure of average efficiency in several hundred traits by means of only seven or eight tests. A not uncommon belief is that such a measure is afforded by most of our scales for the measurement of intelligence. Undoubtedly intelligence tests have their limitations, but improvement in them is to be sought, not by an unlimited increase in the number of tests, but rather by a better selection of tests.

The problem of selecting a small set of tests for measuring average ability in a long list of tests is simply a problem in multiple correlation. Undoubtedly the systematic solution of such a problem would be very laborious. Quite certainly no test battery now in use can be said with assurance to give the highest obtainable multiple correlation with a criterion consisting of average score in a long list of specified tests. The very possibility of such a battery, however, shows how far from necessary it is that a trait should be unitary in order to be measurable. To obtain the data necessary for the construction of such batteries, is, in the writer's opinion, one of the most important tasks confronting the science of psychology. There exist hundreds and hundreds of tests. Now if only one

hundred well selected tests were given to one large group of subjects, and then the correlation of each with all the others calculated, the data obtained would enable one to pick out the one test and the best set (or sets) of 2, 3, 4, 6, 8, etc., for the purpose of measuring average ability in the entire group of tests, or in any section of this group. Such an investigation would be a stupendous undertaking—probably more elaborate than any psychological research yet conducted—but the data not only would have value for the purposes already mentioned, but should, through proper analysis, for the first time yield an adequate scientific basis for the classification of multitudinous mental activities. The results obtained might very well justify the adoption of new technical, classificatory rubrics, definable in terms of test-performance, in place of such terms as attention, character, intelligence, imagination, emotion, reasoning, etc., in the definition of which hardly any two psychologists are in entire agreement. The data from such an investigation would be of great value also in the study of such problems as are raised by Spearman's two-factor theory.

Another topic demanding thorough consideration is the use of ratings as a criterion with which to check the results of mental tests. An awareness of the limitations to the value of ratings, frequently set forth by psychologists, has been expressed elsewhere by the writer.[8] The validity of ratings, as measures of a trait named and defined as the experimenter sees fit to define it, cannot be assumed. There are several points, however, which should be kept in mind. In the first place, in so far as the ratings are simply a matter of chance, or based upon ignorance, the correlations obtained between the ratings of different raters would tend towards zero, and the amalgamated ratings of several judges would tend to show no significant correlation with the scores yielded by any test. It appears not unreasonable, then, to look upon a significant coefficient of correlation between ratings and test scores as evidence, strong in proportion to the magnitude of the obtained coefficient, that the test is measuring something not too remote from what it was proposed to measure—a vague something which the investigator yet without a measure would probably be able neither to define nor to rate much better than any other equally intelligent and experienced person.

A second point to be remembered is that the technique of partial correlation enables one to determine to what extent a correlation

[8] Woodrow, H., 'A Picture-Preference Character Test.' *J. Educ. Psychol.*, 1926, 17, p. 528.

between ratings and test scores is affected by differences between the subjects in any particular measurable factor, such as intelligence or scholarship. As a consequence, it is often feasible to ascertain whether an obtained coefficient represents spurious correlation—any possible sources of which should of course always be carefully considered.

Lastly, in connection with the use of ratings, it is important to emphasize that the validity of ratings, or of any other criterion, is a matter of the definition of the trait which one desires to measure. If one starts with a theoretical definition of a trait, he can never prove that he has an accurate measure of the trait as defined. By a theoretical definition is here meant such a definition as that of intelligence as the 'capacity to learn,' 'the ability of an individual to adjust his thinking to new requirements,' or most of the definitions given some years ago by a number of American psychologists contributing to a symposium.[9] The investigator may find a criterion which to him seems satisfactory, but he can not prove that his criterion correlates perfectly with a true measure of the trait as he would theoretically define it.

A trait may be defined, however, by two other sorts of definitions, which, while perhaps also in a sense theoretical, are clearly enough of a different type from those which are here intended by that term. One of these consists of definitions in terms of the criterion; the other, of definitions in terms of test-performance. All definitions are arbitrary, and if one wishes to use ratings as a tentative, experimental criterion of character, and to define character in terms of this criterion, to the effect that character is that set of traits which determines a person's standing in the amalgamated ratings of a given set of judges, who can deny him the privilege? His criterion then immediately becomes a perfectly valid one, for his definition is in terms of his criterion. Naturally, there will be those who, finding such a rudely arbitrary criterion at variance with either their own theoretical definition or some other criterion, will attach no value at all to such a definition. The temporary use of such definitions, simply as working hypotheses, may, however, by orienting the search for tests, lead to more objective definitions in terms of test-performance, definitions which may command more general assent. A criterion is not a necessity, nor can it ever be proved to accord perfectly with a theoretical definition. It is simply a device for

[9] 'Intelligence and its Measurement: A Symposium,' *J. Educ. Psychol.*, 1926, 12, 123–147 and 195–212.

helping an investigator to decide whether his test will prove useful along the intended lines and whether he is ready to trust his test to the final arbiter of its value, its usefulness in actual practice.

Some of the specific criticisms of character ratings, made by Witty and Lehman, seem to be greatly exaggerated. They write that teachers' ratings show "simply that high academic attainment influences favorably teacher ratings in every other regard." [10] The truth of this statement, so far as it is related to character ratings, may easily be determined by correlations between character ratings and 'academic attainment.' Witty and Lehman give no such data, and the data which the writer happens to have on this point are limited to achievement scores furnished by two tests—the Monroe Reading Test and the Minneapolis Arithmetic Test. These scores were obtained in a study of overstatement in third grade children,[11] in the course of which character rankings were secured. The achievement scores were changed into sigma scores and then combined so as to obtain a single achievement score for each child. These achievement scores were then correlated with the rankings in general character. For the two schools in which these scores were available, the correlations obtained were +.32 and +.42. The average of these two correlations is +.37, a coefficient which is certainly far too low to indicate that the teachers' rankings *simply* reflect scholastic attainments. That they do so to a certain degree does not necessarily invalidate them as ratings of character; for the determination of the extent to which school attainment is a matter of character has not yet been made, and cannot be made without the use of adequate character tests. There is, however, considerable evidence that traits of character are important factors in achievement—in school as well as out of school. In the case of the four-year-old nursery-school children used by Woodrow and Bemmels, it is hard to see just what is the meaning of the criticism that teachers' ratings simply reflect 'academic attainment.' It may well be doubted whether the activities of these little tots are any more 'academic' than those of children of the same age who are not in school. Their day's program is mainly composed of play, but includes luncheon and a nap, and appears to afford a sampling of all the activities that compose the life of such children in any good home. It should be added that it is probably feasible to test certain

[10] Witty, P. A., and Lehman, H. C., 'The So-Called "General Character Test",' PSYCHOL. REV., 1927, 34, 410.

[11] Woodrow, H., and Bemmels, V., 'Overstatement in Third Grade Children,' *J. Appl. Psychol.*, in press, 1928.

character traits in a way which would eliminate the effects of differences in the amount of knowledge possessed by the subjects. In the case of the overstatement test, this has apparently been done by scoring the test by means of the formula devised by Goodenough.[12]

Another criticism against teachers' character ratings is that they are primarily dependent upon whether the child is highly gifted or not, that is, that they are determined by intelligence. That this is not necessarily the case, however, is shown rather decisively by data obtained from seven third grades scattered over the cities of St. Paul and Minneapolis, which yielded correlations between mental age and amalgamated teachers' character rankings which varied from −.20 to +.49, and which averaged only +.25.[13] That there should be some correlation, and at times a fairly high one, between character ratings and intelligence scores does not of itself indicate any lack of validity of the character ratings. For what evidence have we that character and intelligence are not correlated to a certain extent?

That the overstatement test itself is not primarily an intelligence test is readily seen from the fact that, with the third grade children mentioned above, the correlation between overstatement and intelligence scores was only +.30. It may be noted, also, that while the correlations between overstatement scores and teachers' character rankings averaged +.50, the partial correlations, with intelligence score constant, averaged almost the same, namely, +.48.[14]

Witty and Lehman report two bits of evidence to substantiate the view that teachers tend to exaggerate the presence of desirable traits in gifted children. The first line of evidence cited is concerned with teachers' estimates of play. It was first ascertained that, according to the children's own testimony, a group of 50 gifted children engaged in exactly the same number of play activities as a group of normal children. According to the teachers' estimates, however, 45 per cent. of the gifted children played more than an average amount. There is thus indicated a discrepancy between the teachers' estimates and the children's testimony. No data are given, however, which establish what is the cause of this discrepancy; and in the absence of such data, one could conceivably advance numerous explanatory hypotheses other than that the teachers' ratings were primarily determined by the intelligence of the children.

[12] 'Genetic Studies of Genius,' edited by L. M. Terman, 1925, Vol. 1, p. 505.

[13] Woodrow, H., and Bemmels, V., *op. cit.*

[14] In pre-school children, the correlations obtained between overstatement and amalgamated rankings, were, for the five-year-olds, +.56, (not +.86, as stated by Witty and Lehman) and for the four-year-olds, +.43. Calculated with mental age constant, these correlations became +.36 and +.48.

It is conceivable, for example, that the children's testimony may have been obtained in response to different instructions, or differently understood instructions, from those given the teachers; or, since the teachers' estimates were necessarily based upon that very limited portion of the children's play which came under their observation, whereas the children's testimony was based largely on their memory of activities they had engaged in away from school, out of sight of the teacher, that the children and the teachers were really rating different things. If, however, by any chance, the children and the teachers rated different things, a discrepancy in their ratings can be taken as proof of any general trends of error in the estimates of either.

The teachers also rated the gifted children's amenability to discipline. Only 2 per cent of the children were stated not to respond well to discipline. As proof that the teachers' ratings in this matter were influenced by the children's scholarship and intelligence, it is stated that "Observation of these children over a period of three years has convinced the writers that several are decidedly difficult to govern." Apparently we have here a discrepancy between the ratings of the children's teachers and those made by other observers, namely, Witty and Lehman. It is not questioned that the opinions of the latter are the more trustworthy. But the fact that through observations extending over three years 'several' children were found who were unamenable to discipline, though not so rated by their teachers, does not prove that the teachers' errors were brought about either by the high intelligence or the academic achievement of these children. It would first be necessary to show that the increase, beyond the teachers' estimates, in the number of unruly children discovered through three years of observation, was greater in the case of the group of gifted children than would have been the case in a group of non-gifted children.

While certain weaknesses in the data of Witty and Lehman have been mentioned, the writer does not deny the possibility of interpreting their data in conformity with the *assumption* that the teachers' ratings were influenced to a considerable degree by the pupil's intelligence and school attainments. The writer is ready to believe that teachers' judgments of character are not infrequently biased by the intelligence and by the school attainments of the children, and he has at times obtained data which he, presumably like Witty and Lehman, would interpret as reflecting this bias.[18]

[18] See Elonen, A., and Woodrow, H., 'Group Tests of Psychopathic Tendencies in Children,' *J. Abn. & Soc. Psychol.*, in press, 1928.

If, however, it appears reasonable to suppose that teachers' judgments of character may be biased by the intelligence and scholarship of the pupils, the investigator who wishes to measure general character or any of its component traits should not feel that this fact constitutes an insuperable difficulty. He should rather rejoice that the factors most liable to distort teacher's ratings are readily measurable with a fair degree of dependability by means of intelligence and achievement tests. With the aid of these tests and the technique of partial correlation, he may determine the effect, upon correlations between ratings and character test scores, exerted by intelligence and scholarship, and—if past findings continue to be verified—may discover that, even with the influence of these factors entirely eliminated, his correlations remain high enough to promise usefulness on the part of his test.

Many criticisms of character testing other than those discussed above are made by Witty and Lehman, but those which have been considered appear to be the more important ones. Such statements as "Another difficulty encountered in measuring character in the laboratory is the fact that a human being is a constantly changing organism" appear rather trite. The same difficulty is met in measuring reaction-time or the extent of the knee-jerk. It has not proved insurmountable. It is only because we are able to measure reaction-time and the knee-jerk that we know so definitely how much they do vary. Once given a good set of tests of any trait, then it becomes an interesting research to determine the variation in score with differences in 'physiological condition,' 'immediate back-ground of experience' and, of particular interest in the case of character and its component traits, the variation resulting from change in age and from various kinds of 'character training.' Whether the tests are tests simply of class-room habits can probably best be determined by comparing the correlations obtained when the ratings are based upon class-room behavior with those obtained when they are based upon behavior out of school. Voelker, it will be remembered, made considerable use of ratings of boy scout leaders, based upon out-of-school behavior; Raubenheimer used what may be termed court judges' ratings, *i.e.*, the presence of a child in a penal institution—ratings determined presumably by out-of-school behavior, Cady used teachers' ratings. The findings of all three, however, appear to be in fair harmony, and do not indicate that any one of these three criteria is especially superior to the others.

The argument of Witty and Lehman that ethical discrimination is not a test of character, that is, that knowledge of the difference

between right and wrong does not ensure the doing of the right, is merely a reiteration of what has often been said before, and, so far as the writer is aware, is generally conceded. The question of moral imbecility likewise appears to be one which it is unnecessary to raise in connection with the technique of character measurement. This concept has been sufficiently criticised by numerous writers and should be allowed to die its natural death.

Another difficulty, particularly stressed by Witty and Lehman in connection with honesty tests, is that the subject's performance will vary with the strength of temptation, depending upon such variables as the individual's needs at the moment and the liability of detection. This relationship between incentives and acts must certainly be given careful consideration, but whether it is a difficulty or a help in the devising of tests of character may depend upon the skill of the experimenter. It certainly affords a clue to the measurement of honesty, for it suggests that honesty tests may be constructed by affording various degrees of temptation to a subject, with the object of determining what degree of temptation is required before he succumbs. May and Hartshorne appear to have utilized an analogous procedure in providing a series of tests affording widely different opportunities of cheating, and thus measuring the tendency to cheat in terms of the degree of opportunity required to elicit its manifestation. That a man who does not cheat or steal under test conditions may do so under other conditions does not necessarily invalidate an honesty test. Honesty tests aim merely at determining the relative likelihood of honest conduct on the part of different individuals placed in the same situation. They should be criticised neither because they do not enable one to predict what a man will do under unspecified conditions, nor because they do not establish the conditions under which a dishonest act may be committed by a man who, under specified conditions, shows a high degree of honesty.

There also exists the difficulty due to what has been termed 'low trait consistency.' This expression is used to designate the fact that a low correlation may exist between tests which, so far as may be judged from subjective analysis, constitute tests of very similar traits. Thus, the speed with which one set of questions is answered may not correlate very highly with the speed with which a different set is answered, in spite of the fact that each set of questions may be labelled 'speed of decision test.' The low correlations not infrequently obtained in such instances indicate that each test entails a performance of great complexity. As a result of this complexity,

two test-performances, though possessing some common feature causing them to be classified under a single rubric, may yet differ in certain features of more consequence than the one determining their classification. This condition is in no wise peculiar to character traits. The writer found that six memory tests (*i.e.*, tests ordinarily so labelled), given to college sophomores, showed correlations which, after being raised by the correction for attenuation, averaged only +.38.[16] The truth is that any trait which psychologists have yet measured shows some degree of trait inconsistency. It should be noted, however, that while the various particular tests of apparently similar performances may often show a rather low correlation, the average score for a number of such tests may show a relatively high correlation with the average score in a number of others. It may possibly be, therefore, if one defines the ability being measured as the average ability in many tests, that a battery of tests will yield a practically satisfactory measure of the trait as defined.

Finally, however small the interrelationships of a set of tests may be, there is no *logical* difficulty in giving a single name to the average score yielded by the entire set. What performances shall be grouped together for the purpose of measurement is not primarily a question of how closely these performances correlate with each other, but simply a question of the usefulness of the obtained measure.

It is likely that much of the alarm felt in certain quarters over the attempt of psychologists to measure such traits as intelligence, character, honesty, etc., would immediately vanish, once the proposed tests were designated by unfamiliar, non-traditional names; the traits measured defined in terms of test-performance, without futile attempts to prove that the tests measure traits as defined in general, theoretical terms; the problem of the measurement of individual differences clearly distinguished from the determination of the factors causing the differences; and the values claimed for the tests limited to those established through their use. And if, until psychologists are ready to take these drastic steps, conventional, traditional terms continue to be employed in referring to tests, it should be realized that these terms are simply vague, tentative labels, which should not be taken too seriously.

HERBERT WOODROW

UNIVERSITY OF ILLINOIS

[16] 'The Effect of Type of Training upon Transference,' *J. Educ. Psychol.*, 1927, 18, 157–172.

[MS. received March 13, 1928]

DIRECTORY OF

Psychological Review Publications

Original contributions and discussions intended for the Psychological Review should be addressed to

Professor Howard C. Warren, Editor Psychological Review,
Princeton University, Princeton, N. J.

Original contributions and discussions intended for the Journal of Experimental Psychology should be addressed to

Professor Madison Bentley, Editor Journal of Experimental Psychology,
University of Illinois, Urbana, Illinois.

Contributions intended for the Psychological Monographs should be addressed to

Professor Raymond Dodge, Editor Psychological Monographs,
Kent Hall, Yale University, New Haven, Conn.

Reviews of books and articles intended for the Psychological Bulletin, announcements and notes of current interest, and *books offered for review* should be sent to

Professor S. W. Fernberger, Editor Psychological Bulletin,
University of Pennsylvania, Philadelphia, Pa.

Titles and reprints intended for the Psychological Index should be sent to

Professor Walter S. Hunter, Editor Psychological Index,
Clark University, Worcester, Mass

All business communications should be addressed to
Psychological Review Company, Princeton, N. J.

The Psychological Review
is indexed in the
Reader's Guide to Periodical Literature
to be found in most public and
college libraries

DIRECTORY OF
American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.; Cornell University. Subscription $6.50. 624 pages ann. Ed. by M. F. Washburn, K. M. Dallenbach, Madison Bentley and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

The Pedagogical Seminary and Journal of Genetic Psychology—Worcester, Mass.; Clark University Press. Subscription $7.00. 700 pages ann. Ed. by Carl Murchison. Quarterly. Child behavior, differential and genetic psychology. Founded 1891.

Psychological Review—Princeton, N. J.; Psychological Review Company. Subscription $5.50. 480 pages annually. Bi-monthly. General. Founded 1894. Edited by Howard C. Warren.

Psychological Bulletin—Princeton, N. J.; Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literature. Monthly. Founded 1904. Edited by Samuel W. Fernberger.

Psychological Monographs—Princeton, N. J.; Psychological Review Company. Subscription $6.00 per vol. 500 pp. Founded 1895. Ed. by Shepherd I. Franz. Published without fixed dates, each issue one or more researches.

Psychological Index—Princeton, N. J.; Psychological Review Company. Subscription $4.00. 300–400 pp. Founded 1895. Edited by W. S. Hunter. An annual bibliography of psychological literature.

Journal of Philosophy—New York; 515 W. 116th Street. Subscription $4. 728 pages per volume. Founded 1904. Bi-weekly. Edited by F. J. E. Woodbridge, Wendell T. Bush and H. W. Schneider

Archives of Psychology—Columbia University P. O., New York City. Subscription $6. 500 pp. per vol. Founded 1906. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study.

Journal of Abnormal Psychology and Social Psychology—Albany, N. Y. Subscription $5. Boyd Printing Co. Ed. by Morton Prince. In cooperation with Henry T. Moore. Quarterly. 432 pages ann. Founded 1906. Abnormal and social.

Psychological Clinic—Philadelphia; Psychological Clinic Press. Subscription $2.50. 288 pages. Ed. by Lightner Witmer. Founded 1907. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene.

Training School Bulletin—Vineland, N. J.; The Training School. Subscription $1. 160 pages ann. Ed. by E. R. Johnstone. Founded 1904. Monthly (10 numbers). Psychology and training of defectives.

Comparative Psychology Monographs—Baltimore; Williams & Wilkins Co. Subscription $5. 500 pages per volume. Edited by W. S. Hunter. Published without fixed dates, each number a single research.

Psychoanalytic Review—Washington, D. C.; 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Founded 1913. Ed. by W. A. White and S. E. Jelliffe.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Founded 1916. Bi-monthly. Ed. by Madison Bentley.

Journal of Applied Psychology—Baltimore, Md.; Williams & Wilkins Co. Subscription $4. 400 pages annually. Founded 1917. Quarterly. Edited by James P. Porter and William F. Book.

Journal of Comparative Psychology—Baltimore; Williams & Wilkins Company. Subscription $5. 500 pages annually. Founded 1921. Bi-monthly. Edited by Knight Dunlap and Robert M. Yerkes.

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Ed. by Carl Murchison. Monthly. Each number one complete research. Child behavior, differential and genetic psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J. Edited by W. S. Hunter. Subscription $6.00. Monthly. 700 pages annually. Founded 1927.

Journal of General Psychology—Worcester, Mass. Clark University Press. Subscription $7.00. 600–700 pages annually. Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

PR.